FARMACOLOGIA
FÍSICA
GENÈTICA
DRET
GEOLOGIA
INFORMÀTICA
FILOLOGIA
VETERINÀRIA
PEDIATRIA
PERIODISME
BIOLOGIA
SOCIOLOGIA
BOTÀNICA
HISTÒRIA
ECONOMIA
MATEMÀTIQUES
MEDICINA
ART
FILOSOFIA
GEOGRAFIA
PEDAGOGIA
QUÍMICA
ELECTRÒNICA
MÚSICA
BIOTECNOLOGIA
ANTROPOLOGIA
ZOOLOGIA
PSICOLOGIA
PSIQUIATRIA

Universitat Autònoma de Barcelona
Servei de Publicacions

Departament d'Economia
i d'Història Econòmica

Ferran Sancho
Xavier Vilà

100 ejercicios resueltos de estadística básica para economía y empresa

col·lecció
MATERIALS 223

Ferran Sancho
Xavier Vilà

100 ejercicios resueltos de estadística básica para economía y empresa

Departament d'Economia i d'Història Econòmica

Universitat Autònoma de Barcelona
Servei de Publicacions
Bellaterra, 2012

1.ª edición en papel: 2012
1.ª edición corregida: mayo de 2013
1.ª edición digital: 2025

Servicio de Publicaciones
Universitat Autònoma de Barcelona
Edificio A. 08193 Bellaterra (Cerdanyola del Vallès).
Spain
Tel. 93 581 10 22
sp@uab.cat
https://publicacions.uab.cat

ISBN (electrónico): 978-84-10202-90-0

Índice

Prólogo

Tras muchos cursos impartiendo la asignatura de Estadística I en la Facultad de Economía y Empresa de la Universitat Autònoma de Barcelona hemos creído oportuno ofrecer la colección de problemas, junto con sus respectivas soluciones, que se ha ido generando a lo largo del tiempo. En un principio, los problemas se entregaban en listas individuales asociadas y pautadas al desarrollo del curso. Tras una primera experiencia, la lista se conformó en un único documento, eso sí, sin soluciones explícitas.

Con el paso del tiempo hemos ido añadiendo y quitando problemas, depurando la colección, con el objetivo que el nivel de la misma fuese el adecuado al de los estudiantes de la Facultad. Creemos que a pesar de que perduran ciertas diferencias de nivel, y también de extensión, hay una uniformidad suficiente entre los ejercicios y que cualquier alumno de Economía y Empresa debería ser capaz de resolverlos, o como mínimo de empezar a resolverlos, con elevada probabilidad de éxito. No hay en la colección ejercicios imposibles, pero sí hay ejercicios exigentes que precisan un grado de pensamiento profundo y de dedicación y esfuerzo, pero que no requieren capacidad técnica o analítica especial. Los ejercicios son para estudiantes de Economía y Empresa, y por esa razón huimos voluntariamente de los tecnicismos y apostamos por las aplicaciones que pueden ser de interés para los estudiantes de Economía y Empresa. No todos los ejercicios tiene un aire "económico", pero muchos han sido diseñados para que se vea la potencialidad de la estadística en las aplicaciones económicas.

La materia de Estadística en los planes de estudio de los nuevos grados adecuados al plan de Bolonia puede equipararse a la antigua asignatura de Estadística en las licenciaturas de las facultades de Ciencias Económicas y Empresariales. La presente colección corresponde al primer semestre de esta materia en los grados y es aplicable tanto al grado de Economía como a los grados de orientación empresarial, como son los de Administración y Dirección de Empresas, Contabilidad y Finanzas y Empresa y Tecnología. La materia es común a todos los grados, siguiendo el espíritu boloñés. Tras una primera y breve aproximación al tratamiento de datos empíricos, el primer semestre de la materia de Estadística se centra esencialmente en el ámbito del cálculo de probabilidades. Ello incluye una descripción de las reglas básicas de la probabilidad y una introducción extensa a las variables y vectores aleatorios. La inferencia estadística es objeto de estudio en el segundo semestre de la materia.

La paternidad de la colección es amplia, pues ha habido muchos profesores dedicados a la impartición de esta materia y nosotros somos, simplemente, los que con mayor frecuencia hemos asumido la responsabilidad. Si diésemos nombres, lo más probable es que nos dejásemos a alguien por el camino. A pesar de ello, y de ser seguramente injustos, no podemos obviar a Jordi Caballé y Michael Creel, profesores permanentes del departamento responsable de la impartición y colegas de aventura docente en esta materia. Aunque la mayoría de los ejercicios son de cosecha propia hay también un buen número que son "universales" y "atemporales", pues responden a preguntas básicas del cálculo de probabilidades. La colección incluye también un considerable número de ejercicios que aparecieron en las pruebas y exámenes de evaluación. Los hemos distinguido añadiendo un asterisco identificativo (*). De esta manera los estudiantes pueden visualizar cuáles son las dificultades reales a las que se deben enfrentar en el momento de la verdad, cuando se deben aportar los elementos que permitan garantizar que el nivel de conocimientos y de competencias adquiridos son los requeridos.

Bellaterra, junio de 2011

Estadística descriptiva

1. Un país ficticio está compuesto por tres autonomías. La primera (Tacanyuna) tiene dos habitantes cuyas rentas personales son 30 y 25 M (miles de euros). La segunda autonomía (Felicia) tiene tres habitantes con rentas de 45, 62 y 15. La tercera (Andamaria) tiene cinco habitantes con rentas de 38, 86, 43, 65 y 24.

    a. Calcular la renta per cápita de cada autonomía.
    b. Calcular la renta per cápita "promedio" de las autonomías (use la media aritmética simple).
    c. Repetir el apartado anterior usando la media ponderada (piense cuáles son los pesos).
    d. Calcular la renta per cápita de país y compararla con los resultados de b) y c). Comentar.

    a. Calcularemos el promedio. El promedio de un conjunto de observaciones es la suma de los valores del conjunto dividida por el número de observaciones.

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i$$

En cada caso tendremos

$$\bar{x}_T = \frac{30+25}{2} = 27,5$$

$$\bar{x}_F = \frac{45+62+15}{3} = 40,67$$

$$\bar{x}_A = \frac{38+86+43+65+24}{5} = 51,2$$

b. Si calculamos el promedio de las tres rentas tendremos

$$\bar{x}_{CA} = \frac{27{,}5+40{,}67+51{,}2}{3} = 39{,}79$$

c. Tenemos que sumar las rentas de todos los habitantes del país (las tres autonomías) y dividir por el total de habitantes

$$\bar{x}_{per\ capita} = \frac{30+25+45+62+15+38+86+43+65+24}{10} =$$

$$= \frac{55+122+256}{10} = 43{,}3$$

Vemos que la renta per cápita no coincide con el promedio de las rentas per cápita de cada una de las autonomías. Esto se debe a que para calcular la renta per cápita del país a partir de las rentas per cápita de las autonomías estas se tendrían que ponderar por el número de habitantes.

2. (*) La cartera de activos de un inversor está compuesta por dos planes de inversión (planes I y II). El plan I se inició con una aportación de 1.000 euros, tiene dos años de vida y ha obtenido rentabilidades del 10% en el primer año y del 14% en el segundo. El plan II aportó 4.000 euros, tiene un año de vida y ha obtenido una rentabilidad del 9%. Calcular la rentabilidad "promedio" del plan I y estimar la rentabilidad de la cartera de inversiones.

   El plan I compromete 1.000 euros a que las rentabilidades obtenidas se transforman al final de los dos años en 1.254 euros. En efecto, al final del primer año el rendimiento es de $(1{+}0{,}10)\cdot 1.000{=}1.100$ euros. Esta cifra se transforma al final del segundo año en $(1{+}0{,}14)\cdot 1.100{=}1.254$ euros. La rentabilidad "promedio" del plan ha de ser la tasa anual r_I que, aplicada durante dos años a la cifra inicialmente comprometida, genera un rescate final de 1.254. Debe cumplirse pues

$$1.000\cdot(1+r_I)\cdot(1+r_I) = 1.000\cdot(1+r_I)^2 = 1.254$$

   de manera que

$$(1+r_I)^2 = \frac{1.254}{1.000} = 1{,}254$$

   y finalmente

$$r_I = \sqrt{1{,}254} - 1 = 0{,}1198 = 11{,}98\%$$

   Es interesante observar que esta rentabilidad "promedio", calculada usando matemáticas financieras, coincide con la rentabilidad promedio calculada usando la media geométrica a los factores anuales de incremento.

$$(1+r_I) = \sqrt{(1+r_1)\cdot(1+r_2)} = \sqrt{(1{,}10)\cdot(1{,}14)} = \sqrt{1{,}254} = 1{,}1198$$

   Para evaluar la rentabilidad de la cartera hemos de ponderar de alguna manera las rentabilidades "promedio" de los dos planes. Una posibilidad es tener en cuenta que los recursos comprometidos en ambos planes (1.000 y 4.000 euros) son distintos y usarlos como pesos. Procediendo así, la rentabilidad "promedio" de la cartera $\bar{r}$ se podría aproximar usando una media ponderada

$$\bar{r} = \frac{1}{5} \cdot r_I + \frac{4}{5} \cdot r_{II} = (0,2) \cdot (0,1198) + (0,8) \cdot (0,09) = 0,0959 = 9,59\%$$

3. Un aficionado a los coches acaba de adquirir una colección compuesta por:

 1 Ferrari, con un precio de 200.000 €/unidad
 2 Audis, con un precio de 50.000 €/unidad
 4 Seats, con un de precio 15.000 €/unidad
 1 Jaguar, con un precio de 100.000 €/unidad

 a. Calcular el precio "promedio" de un coche (discutir qué medida debe usarse).
 b. Calcular un índice de precios de la compra y compararlo con el anterior.

 a. Se trata de considerar qué media ponderada es adecuada en este caso. Se pregunta cuál es el precio "promedio" de un coche y ello sugiere ponderar por los coches comprados, que en total han sido 8. Podemos usar los siguientes pesos:

$$\omega_1 = \frac{1}{8}$$
$$\omega_2 = \frac{2}{8}$$
$$\omega_3 = \frac{4}{8}$$
$$\omega_4 = \frac{1}{8}$$

de manera que el precio "promedio" de un coche será:

$$\bar{p}_\omega = \sum_{i=1}^{4} \omega_i \cdot p_i = \frac{1}{8} \cdot 200.000 + \frac{2}{8} \cdot 50.000 + \frac{4}{8} \cdot 15.000 + \frac{1}{8} \cdot 100.000 = 57.500$$

 b. Ahora se trata de usar una ponderación que se corresponda con la tradicional ponderación de los índices de precios al consumo. En este caso los pesos se obtienen teniendo en cuenta el gasto dedicado a cada uno de los ítems adquiridos. En otras palabras, para el primer tipo de coche

$$\omega_1 = \frac{n_1 \cdot p_1}{\sum_{i=1}^{4} n_i \cdot p_i} = \frac{1 \cdot 200.000}{1 \cdot 200.000 + 2 \cdot 50.000 + 4 \cdot 15.000 + 1 \cdot 100.000} = 0,435$$

donde n_i indica el número de coches adquiridos de cada una de las marcas. Repitiendo el cálculo para el resto de marcas tendríamos:

$$\omega_2 = \frac{2 \cdot 50.000}{460.000} = 0,217$$

$$\omega_3 = \frac{4 \cdot 15.000}{460.000} = 0,130$$

$$\omega_4 = \frac{1 \cdot 100.000}{460.000} = 0,217$$

Se verifica que los pesos suman la unidad (excepto por error de redondeo). El índice de precios resulta ser en este caso:

$$\overline{p}_\omega = \sum_{i=1}^{4} \omega_i \cdot p_i = 0,435 \cdot 200.000 + 0,217 \cdot 50.000$$
$$+ 0,130 \cdot 15.000 + 0,217 \cdot 100.000 = 121,15$$

Las diferencias de ambos “promedios” son significativas y llamativas. En el primer caso la ponderación se basa en el número de unidades compradas mientras que en el segundo caso se basa en el valor de adquisición de las unidades compradas. Que los resultados sean tan distintos pone de manifiesto la necesidad de pensar con cuidado qué pesos son los más apropiados en cada caso que se analice.

4. Las calificaciones de los alumnos en un examen de Estadística han sido: 6, 4, 4, 3, 6, 10, 1, 0, 2, 6 ,6, 8, 5

   a. Calcular la media aritmética simple, la moda, la mediana y la media geométrica.
   b. Si usted fuese un líder estudiantil, ¿qué medida de centralidad escogería para argumentar la buena “calidad” del grupo?
   c. Si usted fuese el profesor de la materia, ¿qué medida de centralidad escogería para argumentar la pésima “calidad” del grupo?
   d. Si usted fuese un observador imparcial, ¿qué podría decir sobre el nivel del grupo?

   a. La media

$$\overline{x} = \frac{6+4+4+\cdots+8+5}{13} = \frac{61}{13} = 4,69$$

La moda de un conjunto de observaciones es el valor que aparece con mayor frecuencia. Por lo tanto, en esta colección de datos la moda es 6.

Si ordenamos las observaciones en sentido creciente, la mediana es, si no es impar, el valor de la observación que ocupa el lugar central $(\frac{N+1}{2})$. Puesto que $(\frac{N+1}{2})$=7 y el valor de la observación 7, la mediana es 5.

Finalmente, la media geométrica es la raíz *n-ésima* del producto de las observaciones dividida por el número de observaciones

$$\breve{x} = \sqrt[N]{x_1 \cdot x_2 \cdot ... \cdot x_N}$$

En este caso,

$$\breve{x} = \sqrt[13]{6 \cdot 4 \cdot ... \cdot 5} = 0$$

Vemos que la media geométrica queda “desvirtuada” ya que al existir una observación igual a cero hace que esta media geométrica sea cero.

b. La moda, ya que obtenemos la nota más alta con un 6.
c. La media, ya que el resultado es suspenso, con un 4,69.
d. La nota con mayor frecuencia es 6. Los que aprobaron, no lo hicieron con buena nota. Y el promedio está por debajo del aprobado.

5. En una empresa se pagan los siguientes salarios (en euros):

Salarios	Nº empleados
1.200	5
1.300	3
1.400	10
1.500	9
1.600	8
1.700	5
1.800	4

a. Calcular el salario “promedio”.
b. Dilucidar qué grupo de empleados tiene un mayor peso en la formación del salario “promedio” entendido como media ponderada (use dos tipos diferenciados de

pesos).

c. Determinar a qué grupo de empleados corresponde el punto medio de la escala salarial.
d. Uno de los empleados que cobra 1.800 € va a ser trasladado a otra sede y va a ser susstituido por un empleado que recibirá 3.500 €. Verificar si este cambio afecta las respuestas anteriores.

a. En el cálculo del salario "promedio" hemos de tener en cuenta que un mismo salario está compartido por más de un empleado y que el número de empleados es dispar según nivel salarial. Una opción natural es calcular el salario "promedio" usando una media aritmética ponderada en la que los pesos son el número de empleados que comparten un mismo salario sobre el total de empleados. Así, hay 5 empleados que ingresan 1.200 euros mensuales y un total de 44 empleados, de forma que el peso de este nivel salarial sería de 5/44. Otra alternativa sería calcular pesos en función del gasto asociado a cada tipo de salario. En este caso, el gasto en el nivel salarial 1.200 sería de 6.000 euros mientras que el gasto total puede verse que llegaría a 65.900 euros y el peso en este caso sería de 6.000/65.900. La tabla siguiente resume los pesos posibles:

Salario	Empleados	Pesos empleados	Masa salarial	Pesos salarios
1.200	5	0,114	6.000	0,091
1.300	3	0,068	3.900	0,059
1.400	10	0,227	14.000	0,212
1.500	9	0,205	13.500	0,205
1.600	8	0,182	12.800	0,194
1.700	5	0,114	8,500	0,129
1.800	4	0,091	7.200	0,109
Sumas	***44***	***1,000***	***65.900***	***1,000***

Tenemos pues 7 pesos, uno por cada categoría de nivel salarial. Gracias a estos pesos (tercera columna de la tabla) podemos calcular los salarios "promedio" cuando usamos los pesos derivados del número de empleados

$$\bar{x}_w = \sum_{j=1}^{7} w_j \cdot x_j = 1.497$$

Mientras que si usamos los pesos de la masa salarial (quinta columna) obtendríamos:

$$\bar{x}_w = \sum_{j=1}^{7} w_j \cdot x_j = 1.517$$

El salario "promedio" es bastante similar en ambos casos.

b. En ambos casos, el nivel salarial tercero, el que corresponde a 1.400 euros, representa el mayor peso en el cálculo del promedio ponderado.

c. El punto medio de la escala salarial se corresponde con la mediana. En la tabla apreciamos que los niveles salariales están ordenados de forma creciente. Hay 44 empleados, un número par, de forma que la mediana se corresponderá con las posiciones 22 y 23 que pertenecen al cuarto nivel salarial, es decir, con 1.500 euros.

d. El punto medio de la escala salarial medido por la mediana no se verá alterado. En cambio, el salario "promedio" medido por la media ponderada sí que será afectado, pues los pesos se modificarán al haber una nueva clasificación de empelados y niveles salariales. Con pesos basados en el número de empleados el salario "promedio" nuevo pasará a ser $\overline{x}_w = 1.536$, mientras que con pesos basados en la masa salarial será de $\overline{x}_w = 1.612$.

6. El ingreso mensual medio de los empleados agrícolas es de 1.300 euros. El de los empleados no-agrícolas es de 2.000 euros.

   a. Determinar la distribución de empleados que generaría un ingreso medio conjunto de 1.500 euros.
   b. Repetir si el ingreso medio conjunto fuese de 1.800 euros.
   c. Con los datos del apartado anterior, suponer que los empleados no-agrícolas se clasifican en empleados de la industria y empleados de los servicios, con proporciones del 40% y 60% respectivamente. Si el ingreso medio de los empleados de la industria es de 1.900 euros, hallar el de los empleados de servicios.

a. Conocemos el salario medio agrario, el no-agrario y el que corresponde al conjunto de ambos:

$$\overline{x}_A = 1.300$$
$$\overline{x}_{NA} = 2.000$$
$$\overline{x} = 1.500$$

También sabemos que el salario conjunto puede calcularse como una media ponderada del salario agrario y el no agrario, en la que los pesos son las proporciones de empleados en ambas categorías:

$$\overline{x} = \alpha \cdot \overline{x}_A + (1-\alpha) \cdot \overline{x}_{NA}$$

Despejando para la proporción α hallamos que es α=5/7.

b. Se trata de repetir el apartado anterior si ahora $\bar{x} = 1.800$. Hallaríamos que la proporción de empleados agrarios sería α=2/7.

c. El salario medio en la industria es de $\bar{x}_{IND} = 1.900$ con una proporción de empleados industriales del 40% y del 60% en los servicios. El sector no-agrario incluye a la industria y los servicios y sabemos que su salario medio es de 2.000 euros. En estas condiciones tendremos:

$$\bar{x}_{NA} = 0,40 \cdot \bar{x}_{IND} + 0,60 \cdot \bar{x}_{SER}$$

Y de aquí:

$$\bar{x}_{SER} = \frac{\bar{x}_{NA} - 0,40 \cdot \bar{x}_{IND}}{0,60} = \frac{2,000 - 0,40 \cdot 1,900}{0,60} = 2,066,67$$

7. Una empresa de pavimentación de calzadas ha reconstruido 240 metros de calle. La primera mitad se rehizo en 10 días mientras que para la segunda mitad se necesitaron 8 días. El alcalde del pueblo le pregunta al gerente de urbanismo cuál es la productividad "promedio" (metros de calzada por día) de la empresa. Ayude al gerente a responder a esta cuestión.

Vemos que la productividad (espacio pavimentado/tiempo empleado) difiere en las dos mitades de la calzada:

120 metros en 10 días $\rightarrow$ 12 metros/día

120 metros en 8 días $\rightarrow$ 15 metros/día

La productividad en ambas mitades es claramente diferente. Si el gerente cree que la productividad "promedio" es la media aritmética de ambas productividades, le contestará al alcalde que dicha cifra es de 13,5 metros por día (13,5 = (12 + 15)/2). Sin embargo, es inmediato ver que si esta productividad promedio se aplica a las dos mitades se genera una distancia que no corresponde a la efectivamente pavimentada. En efecto,

13,5 metros/día × 18 días = 243 metros

Si el gerente de urbanismo hubiese usado el promedio denominado media armónica, sin embargo, la respuesta hubiese sido otra. En este caso, el promedio de productividad sería:

$$\frac{1}{\frac{1}{2}\cdot\ \frac{1}{12}+\frac{1}{15}} = 13{,}333 \text{ metros/día}$$

Este promedio, aplicado al número total de días empleado en la pavimentación, genera la distancia correcta:

13,333 metros/día × 18 días = 243 metros

8. (*) Una empresa produce coches en tres plantas distintas (A, B y C). En la planta A se producen 500 coches en 20 días, en la B se producen 500 en 25 días y en la C también 500 pero en 30 días.

   a. La dirección de la empresa le pide que le informe de cuál es la productividad "promedio" de sus plantas. Discuta el concepto adecuado de promedio y calcule la información solicitada.
   b. Suponer a continuación que las cifras de producción fueran de 500 en A, 600 en B y 700 en C. Volver a responder a la pregunta de la dirección de la empresa especificando cuál es la contribución de cada una de las plantas a la productividad global de la empresa.

   a. La productividad promedio de la empresa es fácil de calcular pues se producen 1.500 coches en 75 días, es decir, tenemos una productividad promedio de 20 coches por día. Sin embargo, si queremos calcular la productividad de la empresa en función de las productividades de las plantas necesitamos usar el concepto de media armónica. En este caso con productividades de plantas:

A: 500 coches en 20 días: 500/20 = 25

B: 500 coches en 25 días: 500/25 = 20

C: 500 coches en 30 días: 500/30 = 16,667

La productividad de la empresa será:

$$\frac{1}{\frac{1}{3}\cdot\left(\frac{1}{25}+\frac{1}{20}+\frac{1}{16{,}667}\right)} = 20 \text{ coches/día}$$

   b. La productividad promedio es en este caso de 24 (1.800 coches/75 días = 24). Las productividades en las tres plantas son:

A: 500 coches/20 días = 25

B: 600 coches/25 días = 24

C: 700 coches/30 días = 23,333

Si usamos estas tres productividades para obtener el promedio basado en la media armónica obtendríamos:

$$\frac{1}{\frac{1}{3}\cdot\left(\frac{1}{25}+\frac{1}{24}+\frac{1}{23,333}\right)} = 24,092 \text{ coches/día}$$

Cifra que no coincide con la productividad global de la empresa, que hemos visto era de 24. ¿Cuál es el problema aquí? Resulta que en este segundo caso las plantas no están produciendo el mismo volumen de output y es necesario ponderar este hecho. Si usamos la fórmula de la media armónica ponderada:

$$\bar{x}_{\omega}^{a} = \frac{1}{\sum_{i=1}^{n}\omega_i \cdot \frac{1}{x_i}}$$

donde ω_i son pesos no-negativos que miden el tamaño de las plantas y x_i son las productividades en las plantas, encontraremos el resultado correcto:

$$\bar{x}_{\omega}^{a} = \frac{1}{\omega_1\cdot\frac{1}{25}+\omega_2\cdot\frac{1}{24}+\omega_3\cdot\frac{1}{23,333}} =$$

$$= \frac{1}{\frac{500}{1.800}\cdot\frac{1}{25}+\frac{600}{1.800}\cdot\frac{1}{24}+\frac{700}{1.800}\cdot\frac{1}{23,333}} = 24$$

La contribución de cada planta a la productividad de la empresa depende de su tamaño relativo, característica que viene medida por los pesos ω_i.

9. Un recién licenciado está buscando su primer trabajo y ha recibido dos ofertas de empleo por parte de las empresas A y B. El salario medio mensual es el mismo en ambas empresas (1.200 euros), mientras que la desviación estándar en los salarios es de 120

euros en la empresa A y de 60 euros en la B. ¿En qué empresa obtendría el mejor salario inicial? Explicar. ¿En qué empresa obtendría el mejor salario a los diez años de estar empleado? Explicar.

A pesar de compartir el mismo salario mensual medio, la empresa A presenta una escala de salarios más amplia que la empresa B, según se desprende de las desviaciones estándar. Si las retribuciones tienen algo que ver con la experiencia laboral, es de prever que en un principio un empleado (en A o en B) obtendrá un salario inicial cercano a los más bajos de su empresa. Puesto que hay más dispersión salarial en A que en B, el salario inicial en B no sería posiblemente tan bajo como en A. Sin embargo, con el paso de los años y la acumulación de experiencia profesional, un empleado en A podrá llegar a salarios más altos que en B, puesto que el intervalo de los mismos es mayor en A que en B. En definitiva, en el inicio de su carrera B es posiblemente mejor que A, pero con experiencia acumulada A es posiblemente mejor que B.

10. (*) Una empresa está interesada en saber la variabilidad de sus salarios y a qué puede deberse tal variabilidad. La plantilla de la empresa está compuesta en su integridad por cuatro operarios, dos administrativos y dos técnicos. Las retribuciones mensuales (en euros) son:

operario 1: 1.000
operario 2: 1.100
operario 3: 900
operario 4: 1.000
administrativo 1: 700
administrativo 2: 1.300
técnico 1: 1.400
técnico 2: 1.800

a. Calcular las medidas de centralidad y dispersión habituales para la plantilla de la empresa.
b. Catalogar adecuadamente la plantilla y obtener la descomposición de la varianza en intervarianza e intravarianza. Explicar qué fuente de variabilidad es más importante.

a. Lo primero que vamos a hacer es reescalar los datos para evitar que en las operaciones aparezcan números excesivamente grandes. Expresaremos pues las cifras en miles de euros. Tendremos:

Media: $\frac{1+1,1+0,9+1+0,7+1,3+1,4+1,8}{8}=1,150$

Moda: 1 (es el valor que más se repite)

Mediana: si ordenamos las observaciones de menos a mayor tenemos 0,7; 0,9; 1; 1; 1,1; 1,3; 1,4; 1,8.

Por lo tanto la mediana está entre la cuarta y la quinta observación,

$$\frac{1+1,1}{2}=1,05.$$

Varianza: si usamos el símbolo σ^2 para representar la varianza de una colección de datos:

$$\sigma^2 = \frac{(1\text{-}1,15)^2+(1,1\text{-}1,150^2+\cdots+(1,8\text{-}1,15)^2}{8}=0,1025$$

Desviación estándar: $\sqrt{\sigma^2}=\sqrt{0,1025}=0,3202$

b. La colección de observaciones puede catalogarse distinguiendo 3 grupos: operarios, administrativos y técnicos. La variabilidad del conjunto de todas las observaciones puede atribuirse en parte a las diferencias de retribución que existen entre los tres grupos y en parte a las diferencias que se dan dentro de cada grupo. La primera de estas fuentes de variabilidad se denomina intervarianza, mientras que la segunda se denomina intravarianza. Por definición la intervarianza se mide como la varianza de las medias de los grupos mientras que la intravarianza se mide por la media de las varianzas de los grupos. Usando las fórmulas:

$$\sigma_E^2 = \frac{1}{n}\cdot\sum_h\left(\overline{x}_h-\overline{x}\right)^2\cdot n_h$$

$$\sigma_D^2 = \frac{1}{n}\cdot\sum_h\sigma_h^2\cdot n_h$$

donde tenemos que

n: número total de observaciones en la colección (n = 8 en nuestro caso)

h: número de grupos en la colección (h = 3)

n_h: número de observaciones dentro del grupo h (3, 2 y 2)

$\overline{x}$: media de la colección

$\overline{x}_h$: media del grupo h

σ_h^2 : varianza del grupo h

Usando los datos suministrados en el problema hallaríamos:

$$\overline{x}_1 = 1 \quad \overline{x}_2 = 1 \quad \overline{x}_3 = 1{,}6$$
$$\sigma_1^2 = 0{,}005 \quad \sigma_2^2 = 0{,}09 \quad \sigma_3^2 = 0{,}04$$

Y de aquí obtendríamos:

$$\sigma_E^2 = 0{,}0675$$
$$\sigma_D^E = 0{,}035$$

Se verifica que la variabilidad total es la suma de la intervarianza y la intravarianza:

$$\sigma^2 = 0{,}1025 = \sigma_E^2 + \sigma_D^2 = 0{,}0675 + 0{,}035$$

Si expresamos esta propiedad en porcentajes:

$$\frac{\sigma_E^2 + \sigma_D^2}{\sigma^2} \cdot 100 = \frac{0{,}0675}{0{,}1025} \cdot 100 + \frac{0{,}035}{0{,}1025} \cdot 100 = 65{,}85\% + 34{,}15\%$$

de forma que podemos concluir que casi el 66% de la variabilidad en los datos puede atribuirse a la variabilidad que existe entre los grupos (intervarianza).

11. Demostrar que las varianzas de una población y de una muestra también se pueden calcular como:

$$s^2 = \frac{1}{N-1} \cdot \sum_{j=1}^{N} x_j^2 - \frac{N}{N-1} \cdot \overline{x}^2$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} x_j^2 - \overline{x}^2$$

donde x_j representa las observaciones y $\overline{x}$ representa la media aritmética.

En el primer caso y partiendo de la definición tenemos:

$$
\begin{aligned}
s^2 &= \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1} = \frac{\sum_{i=1}^{N}(x_i^2 - 2x_i \cdot \overline{x} + \overline{x}^2)}{N-1} = \\
&= \frac{1}{N-1}\sum_{i=1}^{N} x_i^2 + \frac{1}{N-1}\sum_{i=1}^{N}(-2x_i \cdot \overline{x}) + \frac{1}{N-1}\sum_{i=1}^{N}\overline{x}^2 = \\
&= \frac{1}{N-1}\sum_{i=1}^{N} x_i^2 + \frac{1}{N-1}\sum_{i=1}^{N}(-2x_i \cdot \overline{x}) + \frac{N}{N-1}\overline{x}^2 = \\
&= \frac{1}{N-1}\sum_{i=1}^{N} x_i^2 - 2\frac{N}{N-1}\overline{x} \cdot \overline{x} + \frac{N}{N-1}\overline{x}^2 = \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \frac{N}{N-1}\overline{x}^2
\end{aligned}
$$

En el segundo caso tenemos:

$$
\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N} = \frac{\sum_{i=1}^{N}(x_i^2 - 2x_i \cdot \overline{x} + \overline{x}^2)}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 + \frac{1}{N}\sum_{i=1}^{N}(-2x_i \cdot \overline{x}) + \frac{1}{N}\sum_{i=1}^{N}\overline{x}^2 =
$$

$$
= \frac{1}{N}\sum_{i=1}^{N} x_i^2 + \frac{1}{N}\sum_{i=1}^{N}(-2x_i \cdot \overline{x}) + \overline{x}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2\overline{x} \cdot \overline{x} + \overline{x}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{x}^2
$$

12. Se ha compilado de un censo la información sobre el sexo, la edad y el origen geográfico de una población de individuos, así como sus rentas mensuales en euros. Los datos aparecen en la siguiente tabla:

Individuos	Renta mensual
H,A,N	4.000
M,J,S	1.400
H,J,N	3.500
H,J,S	3.200
H,A,N	3.900
M,A,N	2.400
M,J,N	1.600
M,A,S	2.200
H,J,S	3.100
H,J,N	3.300
M,A,S	2.100
M,J,S	1.300
M,A,S	2.150
M,J,N	1.500
H,J,N	3.300
H,A,N	3.900
M,A,S	2.000
M,A,S	2.200

donde M: masculino, H: femenino, A: adulto, J: joven, N: norte y S: sur. Estudie estos datos intentando extraer toda la información estadística que pueda. En particular:

a. Analice la centralidad de la distribución de las observaciones.
b. Analice la variabilidad de la distribución.
c. Estudie qué papel juegan el sexo, la edad y el origen regional en la variabilidad de la renta.

a. Se trata de calcular la media aritmética de las observaciones. Si reescalamos las cifras para que se expresen en miles de euros tendremos:

$$\overline{x} = \frac{1}{18}\sum_{j=1}^{18} x_j = \frac{(4{,}0+1{,}4+3{,}5+\ldots+2{,}0+2{,}2)}{18} = 2{,}613$$

b. Ahora hemos de calcular la variabilidad de la colección de observaciones salariales y lo hacemos usando la varianza:

$$\sigma^2 = \frac{1}{18}\sum_{j=1}^{18}(x_j - \overline{x})^2 = \frac{(4,0-2,613)^2 + (1,4-2,613)^2 + ... + (2,2-2,163)^2}{18} = 0,792$$

c. El siguiente paso consiste en estudiar como los atributos que escriben las observaciones influencian la variabilidad y a este efecto comenzaremos distinguiendo las observaciones en dos grupos por razón de sexo (H, M). Una vez clasificadas las observaciones calculamos la media aritmética y la varianza en cada uno de los dos grupos:

$$\overline{x}_H = 3,525 \quad \overline{x}_M = 1,885$$
$$\sigma_H^2 = 0,112 \quad \sigma_M^2 = 0,140$$

y a partir de aquí podemos calcular la intervarianza y la intravarianza y descomponer el total de variabilidad en estos dos componentes:

$$\sigma_E^2 = \frac{1}{n}\cdot\sum_h\left(\overline{x}_h - \overline{x}\right)^2 \cdot n_h =$$
$$\frac{1}{18}\cdot\left((2,614-3,525)^2\cdot 8 + (2,614-1,885)^2\cdot 10\right) = 0,664$$
$$\sigma_D^2 = \frac{1}{n}\cdot\sum_h \sigma_h^2 \cdot n_h =$$
$$\frac{1}{18}\cdot\left(0,112\cdot 8 + 0,140\cdot 10\right) = 0,128$$

Podemos verificar que la varianza de la colección completa es la suma de estos dos componentes:

$$\sigma^2 = 0,792 = \sigma_E^2 + \sigma_D^2 = 0,664 + 0,128$$

Si expresamos esta relación en porcentajes veremos que prácticamente el 84% de la variabilidad puede atribuirse a las diferencias de salarios entre sexos:

$$\frac{\sigma_E^2+\sigma_D^2}{\sigma^2}\cdot 100=\frac{0,664}{0,792}\cdot 100+\frac{0,128}{0,792}\cdot 100=83,83\%+16,17\%$$

13. (*) La siguiente tabla presenta información sobre los resultados de un examen de estadística (variable X) y las horas por semana dedicadas al estudio de esta asignatura durante el curso (variable Y) referentes a un grupo de 107 estudiantes.

	menos de 10	***entre 10 y 20***	***más de 20***
Excelente	0	2	2
Notable	2	10	7
Aprobado	14	23	3
Suspendido	37	7	0

En base a la información proporcionada se pide:

a. Obtener las distribuciones de frecuencias marginales, absolutas y relativas, de las dos variables.
b. Si transformamos las calificaciones categóricas en sus equivalentes numéricos tal como se hace en la Facultad de acuerdo con el siguiente baremo:

Excelente = 3, Notable = 2, Aprobado = 1, Suspenso = 0

¿Cuál es la nota promedio de este grupo de estudiantes? ¿Cuál es la desviación estándar? ¿Cuál es la nota mediana?
c. De entre los estudiantes que estudiaron menos de 10 horas, ¿qué porcentaje superó la asignatura?
d. De entre los estudiantes que superaron la asignatura, ¿qué porcentaje estudió menos de 10 horas?
e. ¿Coinciden los porcentajes hallados en los apartados (c.) y (d.)? ¿Por qué?
f. Si transformamos las calificaciones categóricas en numéricas como hemos hecho en el apartado (b.), ¿qué medida mide mejor la dispersión, la desviación estándar o el coeficiente de correlación? Justificar la respuesta.

a. Las frecuencias marginales absolutas se obtienen de la suma por filas (variable X, **calificación** en este caso) y la suma por columnas (variable Y, **horas de estudio** en este caso). Una vez tenemos las frecuencias marginales absolutas, las frecuencias marginales relativas se obtienen dividiéndolas por el total de observaciones, 107 en este caso. Así, a partir de la tabla con los datos originales podemos obtener las frecuencias marginales que se piden:

	menos de 10	*entre 10 y 20*	*más de 20*	$n_{i\cdot}$	$f_{i\cdot}$
Excelente	0	2	2	4	0,0374
Notable	2	10	7	19	0,178
Aprobado	14	23	3	40	0,374
Suspenso	37	7	0	44	0,411
$n_{\cdot i}$	53	42	12	107	
$f_{\cdot i}$	0,495	0,393	0,112		1

b. En este caso, podemos sustituir las calificaciones categóricas (o cualitativas) de la tabla anterior por sus equivalentes numéricos (o cuantitativos).

	menos de 10	*entre 10 y 20*	*más de 20*	$n_{i\cdot}$	$f_{i\cdot}$
3	0	2	2	4	0,0374
2	2	10	7	19	0,178
1	14	23	3	40	0,374
0	37	7	0	44	0,411
$n_{\cdot i}$	53	42	12	107	
$f_{\cdot i}$	0,495	0,393	0,112		1

Media: para calcular la media, dado que no tenemos las observaciones originales si no su resumen en frecuencias, usamos la fórmula:

$$\bar{x} = \frac{1}{n} \cdot \sum n_{i\cdot} \cdot x_i$$

Así tenemos

$$\bar{x} = \frac{1}{107}(4 \cdot 3 + 19 \cdot 2 + 40 \cdot 1 + 44 \cdot 0) = \frac{90}{107} = 0{,}841$$

Desviación estándar: para calcular la desviación estándar (que denotaremos por S) calculamos primero la varianza (S^2). De nuevo, dado que no tenemos las observaciones originales sino su resumen en frecuencias, usamos la fórmula:

$$S^2 = \frac{1}{n}\sum n_{i\cdot}(x_i - \bar{x})^2$$

Así tenemos $S^2 = 0{,}713$

Por tanto,

$$S = \sqrt{S^2} = \sqrt{0{,}713} = 0{,}844$$

Mediana: para calcular la mediana calcularemos primero las frecuencias absolutas acumuladas. Es importante recordar que los valores de la variable deben estar en orden creciente. Por tanto, tenemos que reordenar las filas de la tabla, de menor a mayor. Entonces calculamos las frecuencias acumuladas.

	menos de 10	***entre 10 y 20***	***más de 20***	$n_{i\cdot}$	$N_{i\cdot}$
0	37	7	0	44	44
1	14	23	3	40	84
2	2	10	7	19	103
3	0	2	2	4	107

A continuación, dado que tenemos

$$\frac{n}{2} = \frac{107}{2} = 53{,}5$$

vemos que la observación "mediana" corresponde al valor 1 (Aprobado), ya que 53,5 se encuentra entre la frecuencia acumulada de 0 y la frecuencia acumulada de 1.

c. A partir de la tabla con las frecuencias marginales (apartado a.) observamos que:

- Estudiantes que estudiaron menos de 10 horas = 53
- Estudiantes que superaron la asignatura de
  entre los que estudiaron menos de 10 horas = 0 + 2 +14 = 16

Por tanto, el porcentaje que se pide es:

$$\frac{16}{53} = 0{,}3$$

d. A partir de la tabla con las frecuencias marginales (apartado a.) observamos que:

- Estudiantes que superaron la asignatura = 4 + 19 + 40 = 63

- Estudiantes que estudiaron menos de 10 horas de entre los que superaron la asignatura = 0 + 2 +14 = 16

Por tanto, el porcentaje que se pide es:

$$\frac{16}{63} = 0,25$$

e. No, no coinciden. Y no lo hacen porque se trata de conceptos diferentes.

   Primero calculamos la "frecuencia de aprobados condicionada a que se ha estudiado menos de 10 horas" y luego calculamos la "frecuencia de quienes han estudiado menos de 10 horas *condicionada* a haber superado la asignatura".

f. Claramente, la desviación estándar.

   La desviación estándar es una de las "medidas de dispersión", la raíz cuadrada de la varianza. El coeficiente de correlación mide el grado de relación entre <u>dos</u> variables. Aunque en su cálculo intervienen la covarianza y la desviación estándar, su cálculo no aporta ninguna información sobre la dispersión de las observaciones de <u>una</u> variable con respecto a su media.

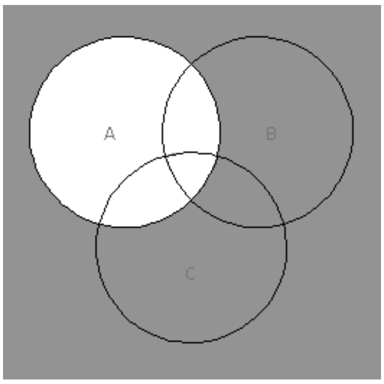
Probabilidad: espacio muestral, sucesos, reglas de la probabilidad

14. Demostrar las siguientes propiedades de conjuntos, donde A, B, y C son subconjuntos de un espacio Ω y donde A^c representa al conjunto complementario del conjunto A

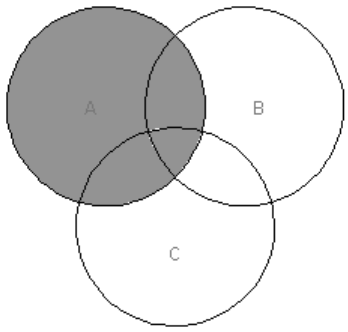
   a. $(A^c)^c=A$
   b. $(A \cup B) \cap C =(A \cap C) \cup (B \cap C)$
   c. $(A \cap B) \cup C =(A \cup C) \cap (B \cup C)$
   d. Leyes de de Morgan: $(A \cup B)^c=A^c \cap B^c$ y $(A \cap B)^c=A^c \cup B^c$
   e. $B =(B \cap A) \cup (B \cap A^c)$

   Para demostrar las igualdades propuestas representaremos, en cada caso, mediante diagramas de Venn el conjunto resultante de las operaciones entre conjuntos que aparecen en la expresión izquierda de la igualdad. La demostración se completa comprobando que la expresión derecha de la igualdad conduce al mismo resultado.

   a. Representamos primero A^c para comprobar luego que su complementario, $(A^c)^c$, es A

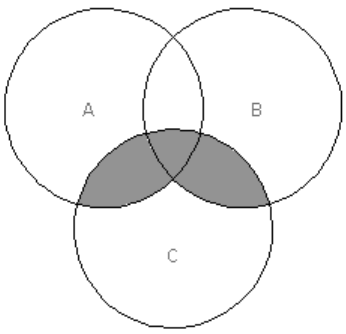


A^c

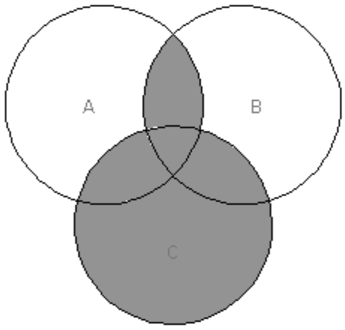


$(A^c)^c=A$

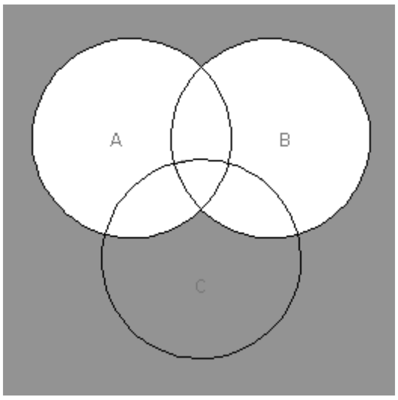
   b. Ambos lados de la igualdad corresponden al conjunto sombreado en el gráfico siguiente:



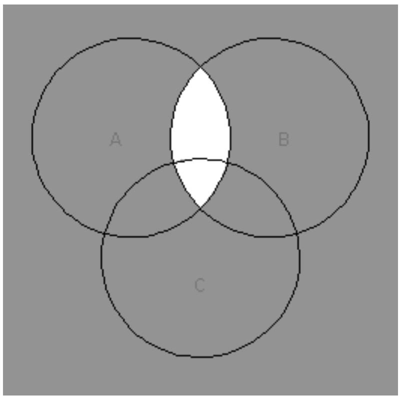
c. Ambos lados de la igualdad corresponden al conjunto sombreado en el gráfico siguiente:



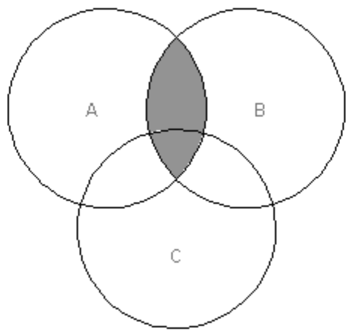
d. Para la primera ley de Morgan tenemos,



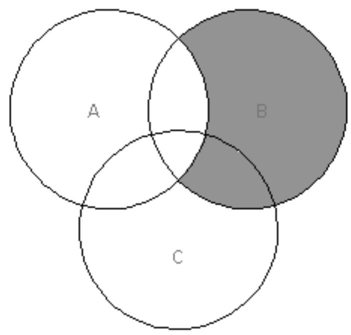
Para la segunda ley de Morgan tenemos,



e. En este caso representamos por separado los dos componentes de la expresión derecha, $\left(B \cap A\right)$ y $\left(B \cap A^c\right)$. Se comprueba inmediatamente que la expresión izquierda, B, coincide con la *unión* de estos dos componentes.



$(B \cap A)$



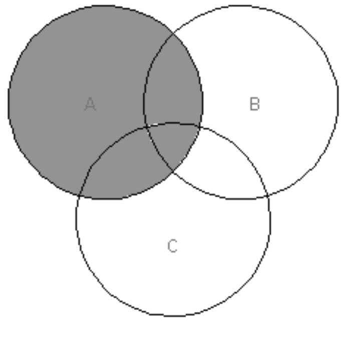
$(B \cap A^c)$

15. Representar gráficamente mediante diagramas de Venn los siguientes supuestos:

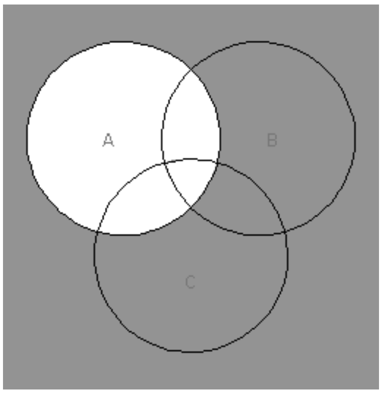
a. El suceso *A* ocurre.
b. El suceso *A* no ocurre.
c. Ocurren los sucesos *A* y *B*.
d. Ocurren *A* o *B*.
e. Ocurren *A* o *B*, pero no ambos.
f. Ocurre *A* pero no ocurre *B*.
g. No ocurren ni *A* ni *B*.
h. Ocurren *A*, *B*, y *C*.
i. Ocurren *A* y *B*, pero no *C*.
j. Ocurren *A* o *B*, pero no *C*.
k. No ocurre *A* pero sí ocurren *B* y *C*.
l. No ocurre *A* pero sí ocurren *B* o *C*.

El área sombreada representa, en cada caso, el suceso propuesto:

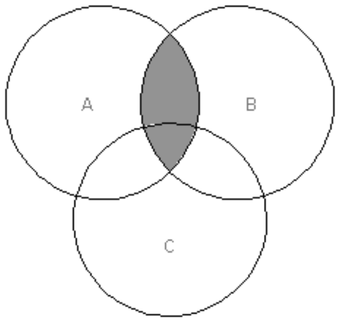
a.



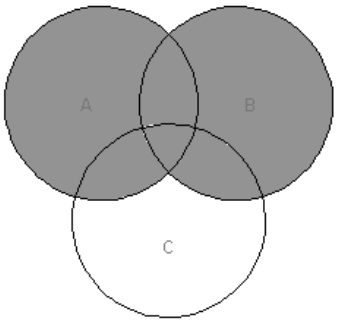
b.



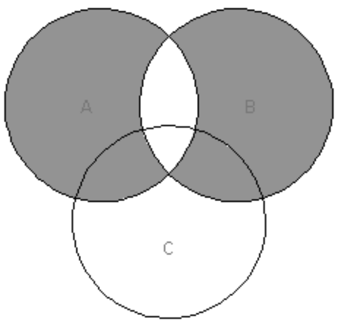
c.



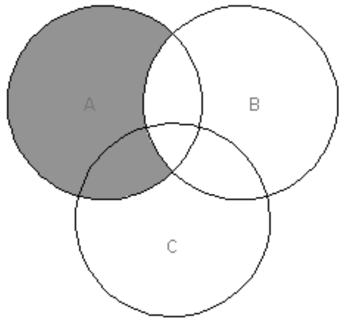
d.



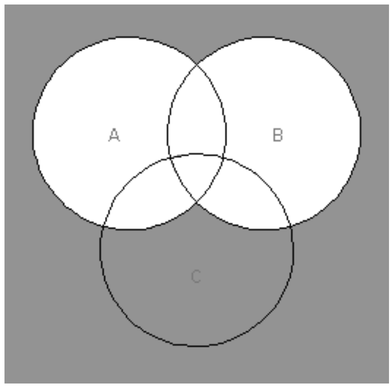
e.



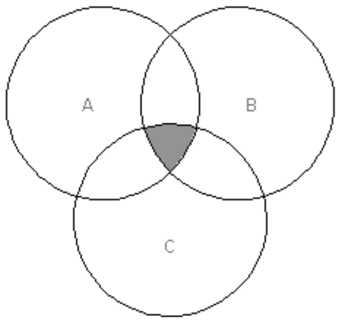
f.



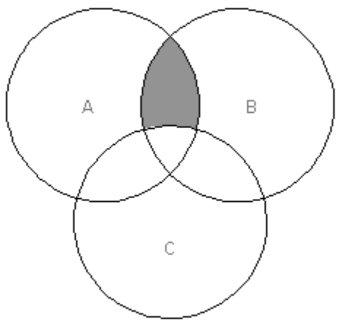
g.



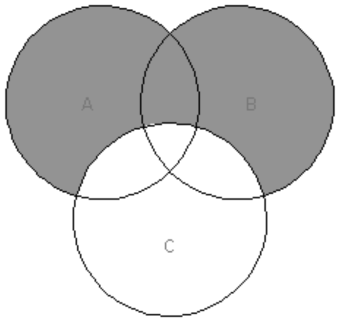
h.



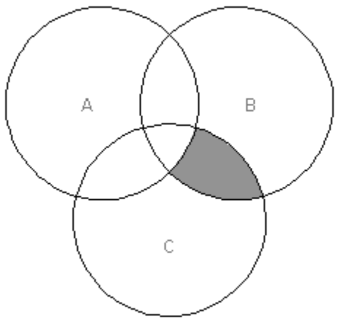
i.



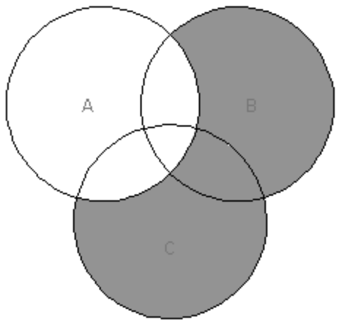
j.



k.



l.



16. Una urna contiene una bola blanca, una bola negra y una verde. Considere los experimentos aleatorios siguientes:

    a. Se extraen secuencialmente dos bolas de la urna. Una vez extraída la primera se anota el color y se vuelve a poner en la urna; a continuación se extrae por segunda vez una bola y se vuelve a anotar el color (este procedimiento se denomina muestreo con reemplazamiento). Construir el espacio muestral Ω .
    b. Este apartado es como el anterior, pero una vez la primera bola ha sido extraída no se devuelve a la urna (este procedimiento se denomina muestreo sin reemplazamiento). Construir el espacio muestral Ω.

c. Determinar las probabilidades de los sucesos simples e_i que componen Ω en el apartado a. utilizando el concepto de probabilidad objetiva clásica (casos favorables sobre casos totales).
d. Lo mismo que en c. para los sucesos simples del apartado b.

a. Representemos el color de cada bola por su inicial: *B*, *N*, *V*. Dado que las bolas se extraen secuencialmente y que se habla de "primera bola" y de "segunda bola" entenderemos que el orden importa. En este caso, dado que hay reposición, tenemos:

$$=\left\{(B,B),(B,N),(B,V),(N,B),(N,N),(N,V),(V,B),(V,N),(V,V)\right\}$$

b. Al no haber reposición una misma bola no puede aparecer dos veces. Por tanto,

$$=\left\{(B,N),(B,V),(N,B),(N,V),(V,B),(V,N)\right\}$$

c. Dado que tenemos 9 resultados (sucesos simples) posibles, todos ellos equiprobables, la probabilidad de cada uno de ellos será:

$$p(e_i)=\frac{1}{9}$$

d. De la misma manera, si ahora hay 6 resultados equiprobables diferentes tendremos:

$$p(e_i)=\frac{1}{6}$$

17. Una urna contiene dos bolas blancas y una bola negra. Considerar el experimento aleatorio de extraer secuencialmente dos bolas de la urna y anotar el color.

a. Construir el espacio muestral Ω si la extracción es con reemplazamiento.
b. Lo mismo si la extracción es sin reemplazamiento.
c. Determinar las probabilidades de los sucesos simples e_i del apartado a) usando el concepto clásico de probabilidad.
d. Lo mismo que en c) para los sucesos simples del apartado b).
e. Calcular la probabilidad del suceso compuesto *A*: "las bolas son del mismo color" para los dos experimentos de extracción.

a. Representamos con B_1 y B_2 las dos bolas blancas, y con N la bola negra. Si hay reemplazamiento el espacio muestral será:

$$\Omega = \{(B_1, B_1), (B_1, B_2), (B_1, N), (B_2, B_1), (B_2, B_2), (B_2, N), (N, B_1), (N, B_2), (N, N)\}$$

b. Sin reemplazamiento tenemos:

$$\Omega = \{(B_1, B_2), (B_1, N), (B_2, B_1), (B_2, N), (N, B_1), (N, B_2)\}$$

c. Dado que tenemos 9 resultados (sucesos simples) posibles, todos ellos equiprobables, la probabilidad de cada uno de ellos será:

$$p(e_i) = \frac{1}{9}$$

d. De la misma manera, si ahora hay 6 resultados equiprobables diferentes tendremos:

$$p(e_i) = \frac{1}{6}$$

e. En el caso con reemplazamiento, hay 5 resultados en los que las dos bolas son del mismo color. Por tanto,

$$p(A) = \frac{5}{9}$$

Si no hay reemplazamiento, solo en 2 resultados de los 6 posibles tenemos las dos bolas del mismo color. Así,

$$p(A) = \frac{2}{6}$$

18. Imaginar un dado ficticio “honesto” que tuviera únicamente tres caras numeradas 1, 2 y 3.

a. Construir el espacio muestral Ω correspondiente al experimento aleatorio repetible de lanzar el dado dos veces y anotar los números que aparecen.
b. Determinar la probabilidad de cada uno de los sucesos simples e_i de Ω.

Considerar a continuación los sucesos compuestos definidos por:

A_1: no aparecen ni el 1 ni el 2.
A_2: aparece al menos un 2.
A_3: la suma de los números que aparecen es menor que 5.

c. Definir estos sucesos en base a los sucesos simples que los componen.
d. Definir los sucesos simples que componen los sucesos:

$$A_1 \cup A_2,\ A_1 \cup A_3,\ A_2 \cup A_3,\ A_1 \cap A_2$$
$$A_1 \cap A_3,\ A_2 \cap A_3,\ A_1 \cup A_2^C,\ A_1 \cup A_3^C,\ A_1 \cup (A_2 \cap A_3)$$

e. Determinar la probabilidad de cada uno de los sucesos compuestos anteriores

a. Representaremos los posibles resultados del experimento de azar por duplas del estilo (1,1). En este caso estaríamos diciendo que al lanzar el dado dos veces aparece un 1 en cada lanzamiento. El espacio muestral será:

$$\Omega = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}$$

b) Puesto que el dado es “honesto”, cada una de las tres caras es equiprobable. Asimismo, y por la misma razón, cada una de las 9 duplas en Ω será un suceso equiprobable. Puesto que el total de probabilidad es siempre 1, ello conduce a concluir que cada dupla tiene probabilidad 1/9.

c) Se trata de identificar los sucesos simples que forman parte de cada uno de los nuevos sucesos. Tendremos:

$$A_1 = \{(3,3)\}$$
$$A_2 = \{(1,2),(2,1),(2,2),(2,3),(3,2)\}$$
$$A_3 = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)\}$$

d) Similar al caso anterior, se trata de usar las propiedades de las operaciones de conjunto para listar los sucesos simples que constituyen cada uno de los sucesos compuestos. A título de ejemplo presentamos algunos de ellos y dejamos el resto para el lector:

$$A_1 \cup A_2 = \{(1,1),(2,1),(2,2),(2,3),(3,2),(3,3)\}$$
$$A_1 \cap A_2 = \varnothing$$
$$A_2 \cap A_3 = \{(1,2),(2,1),(2,2)\}$$

e) Para los ejemplos previos:

$$P(A_1 \cup A_2) = 6/9$$
$$P(A_1 \cap A_2) = 0$$
$$P(A_2 \cap A_3) = 3/9$$

19. (*) Vamos a suponer que se presentan 4 candidatos a las futuras elecciones a la presidencia del Barça (A, B, C y D). Un estadístico profesional ha establecido sin ningún tipo de duda que la probabilidad de ser escogido *president* es directamente proporcional al número de años que el candidato ha participado en las tertulias futboleras (*La Gradería*, *El Rondo*, *Punto Pelota*, etc.). Para los 4 candidatos estas experiencias son, respectivamente, de 2, 3, 4 y 5 años.

   a. Evaluar la probabilidad que tiene cada candidato a ser elegido *president*.
   b. Evaluar la probabilidad que sea elegido presidente uno de los 2 candidatos con menor experiencia tertuliana.

   a. Representaremos por A el suceso "el candidato A ha sido elegido *president*", ídem con B, C y D. Si hacemos que p represente la probabilidad de base, de momento desconocida, que corresponde a 1 año de experiencia en tertulias futboleras, tendremos que por proporcionalidad se ha de cumplir:

$$P(A) = 2 \cdot p$$
$$P(B) = 3 \cdot p$$
$$P(C) = 4 \cdot p$$
$$P(D) = 5 \cdot p$$

Además el espacio muestral es $\Omega = \{A, B, C, D\}$. Sabemos que $P(\Omega) = 1$ y también:

$$P(\ ) = P\{A, B, C, D\} = P(A) + P(B) + P(C) + P(D) = 2 \cdot p + 3 \cdot p + 4 \cdot p + 5 \cdot p$$

y de aquí se encuentra que p= 1/14. Por consiguiente, $P(A)$=2/14, etc.

b. Se trata del suceso $A \cup B$. Por la regla aditiva y dado que son sucesos excluyentes (ambos no pueden ser *president*) tendremos:

$$P(A \cup B) = P(A) + P(B) = 2/14 + 3/14 = 5/14 \approx 36\%$$

20. Los ordenadores han de estar protegidos de los ataques de los virus informáticos. Estamos considerando dos productos antivirus. El primero (A) se sabe que detecta un 50% de los virus, mientras que el segundo (B) solo detecta el 40% de los virus informáticos. Queremos saber cuál sería la efectividad protectora de usar ambos programas a la vez. Y para ello hemos de evaluar el porcentaje de detección si se usan ambos programas (pista: una manera de proceder es partir de calcular los porcentajes mínimo y máximo de efectividad conjunta que corresponden al peor y mejor escenario de efectividad conjunta.)

El peor escenario posible es aquel en el que los programas A y B comparten la detección de una misma base de virus pero A detecta un poco más que B (es decir: $B \subseteq A$). En este caso los virus detectados por ambos programas serían $A \cap B = B$ de forma que $P(A \cap B)$ = $P(B)$=0,40. En el mejor escenario posible A y B detectan virus totalmente distintos, de forma que $A \cap B = \varnothing$ y así $P(A \cap B)$ = 0. El suceso "efectividad en la protección si se usan ambos programas" se puede representar por $A \cup B$. Por la regla aditiva de la probabilidad:

$P(A \cup B)$=$P(A)$ + $P(B)$ – $P(A \cap B)$ = 0,90 (en el mejor escenario)

$P(A \cup B)$=$P(A)$ + $P(B)$ – $P(A \cap B)$ = 0,50 (en el peor escenario).

21. Una caja contiene cuatro monedas, 3 de las cuales son "honestas" y una está trucada (tiene dos caras). Una moneda es escogida al azar de la caja y es lanzada al aire dos veces. Considere el suceso A: "aparecen dos caras". Hallar la probabilidad $p(A)$ del suceso.

Hay dos posibilidades mutuamente excluyentes de observar el suceso. La primera es si la moneda elegida es la "honesta" (llamemos a esta posibilidad E_1, que tiene probabilidad 3/4) y la segunda si la moneda es la trucada (E_2 con probabilidad 1/4). En el primer caso se observarán dos caras con probabilidad 1/4, mientras que en el segundo caso se observarán las dos caras con certeza, o probabilidad 1. Se trata de ponderar ambas vías para llegar al suceso deseado y ello conduce a usar la regla de la probabilidad total

$$P(A) = P(E_1) \cdot P(A/E_1) + P(E_2) \cdot P(A/E_2) = \frac{3}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot 1 = \frac{7}{16}$$

22. Suponer que de una tabla actuarial se sigue que la probabilidad que un individuo de 30 años muera en el transcurso de un año es de 0.1. Tomar un grupo de 10 individuos de esa edad y determinar las probabilidades de los siguientes sucesos:

    a. A_1: ninguno de ellos muera.
    b. A_2: no muera más de uno.
    c. A_3: muera más de uno.
    d. A_4: muera por lo menos uno.

    a. Tenemos 10 individuos, j=1,2,..,10, y cada uno de ellos puede seguir viviendo (suceso V_j) o fallecer durante el periodo en cuestión (suceso M_j). El suceso A_1 puede visualizarse como el suceso "todos los individuos perviven" que podemos suponer son independientes entre si y usar la regla multiplicativa de la probabilidad:

    $$P(A_1) = P(V_1 \cap V_2 \cap ... \cap V_{10}) = P(V_1) \cdot P(V_2) \cdot ... \cdot P(V_{10}) = (0,9)^{10} = 0,349$$

    b. Este suceso corresponde a una situación en la que todos perviven o como mucho muere uno de los individuos. Para el primer caso hemos evaluado la probabilidad en el apartado previo. Para el segundo caso podemos empezar considerando que muere el primero de los individuos. Sería un suceso con probabilidad evaluada por:

    $$P\left(M_1 \cap V_2 \cap V_3 ... \cap V_{10}\right) = P(M_1) \cdot P(V_2) \cdot ... \cdot P(V_{10}) = (0,1) \cdot (0,9)^9 = 0,039$$

    Ahora bien, hay 10 secuencias distintas que tienen la misma probabilidad, en función de que quien muera sea el primero, segundo, etc. Así pues, la probabilidad que muera solo uno de los 10 individuos será:

    $$10 \cdot (0,1) \cdot (0,9)^9 = 0,387$$

    La probabilidad del suceso A_2 incluye esta probabilidad más la de que no muera ninguno, calculada en a), al ser ambos sucesos excluyentes. En definitiva:

    $$P(A_2) = 0,349 + 0,387 = 0,736$$

c. El suceso A_3 puede verse fácilmente que es el complementario del suceso A_2. Por consiguiente:

$$P(A_3) = 1 - P(A_2) = 1 - 0{,}736 = 0{,}264$$

d. El suceso A_4 es el complementario del suceso A_1. Usando de nuevo la regla del complementario hallamos:

$$P(A_4) = 1 - P(A_1) = 1 - 0{,}349 = 0{,}651$$

23. De una encuesta se deduce que el porcentaje de individuos entrevistados que leen el periódico A es del 25% mientras que los que leen el periódico B es del 40%.

    a. Discutir y comentar si estos sucesos son independientes y/o mutuamente excluyentes.
    b. En el caso que fueran independientes, calcular la probabilidad que un individuo lea por lo menos uno de los periódicos.
    c. En el caso que fueran excluyentes, volver a calcular la misma probabilidad.

a. *No tienen por qué ser excluyentes (si leo Sport, ello no excluye que lea El Mundo Deportivo*) ni independientes (si leo *Sport*, la probabilidad de leer *El Mundo Deportivo* es seguramente más alta que si no leo *Sport* al indicar que soy fanático del deporte).

b. Usando las probabilidades suministradas y la regla aditiva, si los sucesos fuesen independientes tendríamos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A) \cdot P(B) = 0{,}55$$

c. Si fuesen excluyentes tendríamos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(\emptyset) = 0{,}65$$

24. Suponer que la probabilidad que un consumidor repita marca en la compra de renovación de un electrodoméstico es p. Hoy está usando la marca A.

    a. Calcular la probabilidad de que en la segunda renovación del electrodoméstico siga comprando la marca A.
    b. Calcular la probabilidad de que en la segunda renovación cambie de marca.
    c. Determinar el porcentaje de clientes fieles, entre los que usan hoy la marca, para

que la empresa productora del electrodoméstico *A* consiga que el porcentaje de segundas renovaciones sea del 80%.

d. Discutir qué vías podría usar la empresa para incentivar la fidelidad de los clientes a su marca.

a. Definamos por R_n el suceso enésima renovación. Dos renovaciones suponen probabilidad:

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2 / R_1) = p \cdot p = p^2$$

siempre que la segunda renovación no dependa de la primera.

b. No renovar es el suceso complementario de renovar. Ahora tendremos:

$$P(R_1 \cap R_2^c) = p \cdot (1 - p)$$

c. Para que el porcentaje de personas que renueven dos veces sea del 80% hace falta que se cumpla $p^2 \geq 0{,}8$ de manera que la probabilidad de renovación habría de ser como mínimo de $p = 0{,}894$.

d. A mayor probabilidad de renovación, mayor fidelidad de los clientes. Una mejora de la calidad percibida del producto, las condiciones de la garantía, o la calidad del servicio postventa servirán para incrementar la probabilidad p.

25. Una empresa de venta de productos por correo considera tres tipos de errores que pueden ocurrir al enviarse un pedido.

A: el artículo que recibe el cliente no es el solicitado.
B: el artículo se extravía y el cliente no lo recibe.
C: el artículo que recibe el cliente sufre desperfectos durante el transporte.

En cualquiera de estos tres casos el cliente está insatisfecho.

a. Dar un argumento para sostener que *A* y *B* son independientes.
b. Ídem con *A* y *C*.
c. Dar un argumento para sostener que *B* y *C* son mutuamente excluyentes.
d. Suponer que las probabilidades de ocurrencia de los errores son $p(A)$=0,02, $p(B)$=0,01 y $p(C)$=0,04. Calcular el porcentaje de clientes insatisfechos con el servicio de la empresa.

a. Si el cliente no recibe nada, no puede recibir un producto distinto.

b. Si hay desperfectos, ellos tienen poco que ver con si el artículo recibido es el solicitado, o no.

c. Si se recibe algo, no puede ser que no se reciba nada.

d. Podemos suponer que si se sufre uno de los tres errores el cliente está insatisfecho. En este caso, el porcentaje de clientes insatisfechos se puede calcular a partir de la probabilidad:

$$P(A \cup B \cup C)$$

Usando la regla aditiva a la unión de tres sucesos, este cálculo requiere evaluar:

$$\begin{aligned} P(A \cup B \cup C) = P(A) + P(B) + P(C) \\ - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ + P(A \cap B \cap C) \end{aligned}$$

Usando las probabilidades dadas y aceptando los argumentos previos en a., b. y c. tendremos:

$$\begin{aligned} &p(A \cap B) = p(A) \cdot p(B) = (0{,}02) \cdot (0{,}01) \\ &p(A \cap C) = p(A) \cdot p(C) = (0{,}02) \cdot (0{,}04) \\ &p(B \cap C) = p(\varnothing) = 0 \\ &p(A \cap B \cap C) = P(B \cap C) \cdot P(A / B \cap C) = 0 \end{aligned}$$

Sustituyendo estos valores hallaríamos el porcentaje de clientes insatisfechos:

$$P(A \cup B \cup C) = 0{,}069 = 6{,}9\%$$

26. El valor de las cotizaciones de las acciones está relacionado con el crecimiento del PIB de la economía, de forma que cuando el PIB aumenta (cae, o se mantiene) el valor de las acciones aumenta (cae, o se mantiene) con determinada probabilidad. Cuando el PIB aumenta, la probabilidad que el valor de las acciones aumente es del 80%, mientras que cuando el PIB se mantiene la probabilidad de aumento de las acciones es del 20% y, finalmente, cuando el PIB cae solo existe un 10% de probabilidades de que las acciones aumenten de valor. Suponga que se espera que el PIB seguirá creciendo en un horizonte a corto plazo con probabilidad 0,6. Con un 0,3 de probabilidad se espera que se mantendrá y también con un 0,1 se espera que caerá. Averiguar la probabilidad que las acciones aumenten de valor en dicho horizonte.

    Vamos a usar la siguiente notación para analizar esta cuestión:

    V: el valor de las cotizaciones aumenta
    A_1: el PIB aumenta, con $P(A_1)=0{,}6$ y además $P(V/A_1)=0{,}8$
    A_2: el PIB se mantiene igual, con $P(A_2)=0{,}3$ y $P(V/A_2)=0{,}2$
    A_3: El PIB cae, con $P(A_3)=0{,}1$ y $P(V/A_3)=0{,}1$

    Las tres situaciones A_1, A_2 y A_3 forman una colección exhaustiva y excluyente de sucesos. Podemos pues usar la regla de la probabilidad total para evaluar la probabilidad que las acciones aumenten de valor:

$$P(V) = P(A_1)\cdot P(V / A_1) + P(A_2)\cdot P(V / A_2) + P(A_3)\cdot P(V / A_3) = 0{,}55$$

27. (*) ***La Guerra de las Galaxias***. La nave estelar Halcón Milenario ha de viajar del planeta Tatooine al planeta Aldebarán. La nave dispone de dos motores de impulso iónico, cada uno de los cuales tiene una probabilidad $p = 0{,}10$ de fallar durante el vuelo. Si fallan los dos motores la nave queda inutilizada y detenida en el espacio. Si falla uno de los motores, sin embargo, la nave continua operativa y funcional. El capitán Han Solo planea tomar la ruta más directa entre ambos planetas.

    a. Evaluar numéricamente la probabilidad que la nave llegue a su destino.
    b. El servicio de espionaje de Darth Vader ha averiguado la ruta que la nave de Han Solo va a tomar y ha enviado cruceros imperiales para capturar al Halcón Milenario. Han Solo, en consecuencia, decide cambiar la ruta y adentrarse en el cinturón de asteroides conocido como Rocky. Se sabe que de cada 100 naves que cruzan el cinturón sin detenerse, solo 10 son destruidas por un asteroide. Una vez cruzado el cinturón, la nave de Han Solo usa una nueva ruta en la que no hay ningún crucero imperial. Evaluar la probabilidad que la nave sea destruida en el tránsito en el cinturón.
    c. Evaluar la probabilidad que la nave llegue a Aldebarán si Han Solo usa la ruta de los asteroides.
    d. Como Darth Vader tiene un magnífico servicio de espionaje galáctico, sabe que Han Solo va a usar la ruta de los asteroides y decide enviar dos cruceros imperiales a la zona de salida del cinturón de asteroides en el que debe aparecer el Halcón Milenario, si es que antes no ha sido destruido por un asteroide. Cada crucero está

equipado con un sensor positrónico de localización capaz de detectar una de cada cuatro naves que circulan por su zona de actuación. Evaluar la probabilidad que Han Solo llegue a Aldebarán.

a. Vamos a introducir la siguiente notación que será de ayuda para encarar la resolución del problema. Llamaremos:

   A: El Halcón Milenario funciona en cualquier momento del vuelo.
   D: La nave es destruida al cruzar el cinturón de asteroides.
   T: la nave es detectada por las fuerzas imperiales.
   B: la nave llega a Aldebarán.
   F: los dos motores del Halcón Milenario fallan.

   La nave llega a su destino (suceso *B*) siempre que funcione (suceso *A*). Y funciona siempre que no fallen los dos motores (suceso *F*). Si los motores iónicos funcionan independientemente, la probabilidad de que ambos fallen será de:

$$P(F) = p \cdot p = 0{,}01$$

   y así

$$P(B) = P(A) = 1 - P(F) = 0{,}99$$

b. La nave puede ser destruida (suceso *D*) tanto si funciona (suceso *A*) como si no funciona (suceso complementario A^c). Además, con probabilidad 0,1 una nave que funciona es destruida en el cinturón, y con certeza lo son las que no funcionan. Usando esta información en la regla de la probabilidad total tenemos:

$$P(D) = P(A) \cdot P(D / A) + P(A^c) \cdot P(D / A^c) = (0{,}99) \cdot (0{,}1) + (0{,}01) \cdot 1 = 0{,}109$$

c. En este caso, la nave llega (*B*) si cruza el cinturón sin ser destruida (D^c) y está funcional (*A*) en el resto del trayecto. En otras palabras $B = A \cap D^c$ de forma que:

$$P(B) = P(A \cap D^c) = P(D^c \cap A) = P(D^c) \cdot P(A / D^c) = P(D^c) \cdot P(A)$$

   puesto que $P(A / D^c) = P(A)$, atendiendo a que una vez la nave no ha sido destruida en el cinturón de asteroides, la probabilidad de seguir funcionando es la misma. En consecuencia:

$$P(B) = P(D^c) \cdot P(A) = (1 - P(D)) \cdot P(A) = (1 - 0{,}109) \cdot 0{,}99 = 0{,}882$$

d. Finalmente, en este apartado, la nave llega a Aldebarán (B) si cruza el cinturón sin ser destruida (D^c), no es detectada (T^c) y está funcional (A). Estos tres sucesos son independientes de forma que usando la regla multiplicativa:

$$P(B) = P(D^c \cap T^c \cap A) = P(D^c) \cdot P(T^c) \cdot P(A)$$

Por otra parte, hay dos cruceros imperiales intentando detectar al Halcón Milenario y cualquiera de ellos puede detectarlo. O sea que $T = T_1 \cup T_2$. Usando la regla aditiva:

$$P(T) = P(T_1 \cup T_2) = P(T_1) + P(T_2) - P(T_1 \cap T_2) = 0{,}25 + 0{,}25 - 0{,}25^2 = 0{,}438$$

De aquí podemos concluir ahora que la probabilidad que el viaje del Halcón Milenario llegue a buen puerto se evaluaría por:

$$P(B) = P(D^c) \cdot P(T^c) \cdot P(A) = (0{,}891) \cdot (0{,}562) \cdot (0{,}99) = 0{,}496$$

28. Una empresa petrolera está inspeccionando una zona para estudiar la existencia de nuevos yacimientos. Los geólogos de la empresa consideran que la probabilidad que el terreno contenga petróleo es del 60%. Sin embargo, al efectuar un análisis del subsuelo a 200 metros de profundidad encuentran que el subsuelo contiene carbonita. La experiencia demuestra que el 20% de los terrenos con yacimientos de petróleo están rodeados por este tipo de subsuelo mientras que el 80% de los terrenos que no tiene petróleo tienen este subsuelo. Dada esta información, averiguar la probabilidad de que el terreno contenga petróleo.

Cuando se hace la exploración se acaba encontrando petróleo o no. Vamos a utilizar la notación A_1: se encuentra petróleo y A_2: no se encuentra. Además B representará el suceso "hay carbonita en el subsuelo". Sabemos:

$P(A_1) = 0{,}6$
$P(B/A_1) = 0{,}2$
$P(B/A_2) = 0{,}8$

Mientras que deseamos calcular la probabilidad que haya petróleo si el subsuelo contiene carbonita. Usando la regla de Bayes y las probabilidades suministradas podemos escribir y hallar:

$$P(A_1 / B) = \frac{P(A_1) \cdot P(B / A_1)}{P(A_1) \cdot P(B / A_1) + P(A_2) \cdot P(B / A_2)} = 0{,}2727 = 27{,}27\%$$

29. (*) La probabilidad que llueva en Alemania mañana (suceso A) es $P(A)$=0,5. La probabilidad que llueva en España mañana (suceso E) es $P(E)$=0,2. La probabilidad que llueva en Alemania y en España mañana es 0,15. Se pide:

    a. Calcular la probabilidad de que llueva en Alemania o en España mañana.
    b. ¿Son los sucesos A y E independientes? Justificar la respuesta.
    c. Suponga que ya es mañana y nos dicen que está lloviendo en Alemania, ¿cuál es la probabilidad de que llueva en España?

    a. Sabemos lo siguiente: $P(A) = 0{,}5$; $P(E) = 0{,}2$; $P(A \cap E) = 0{,}15$

    Simplemente, aplicando la regla aditiva, o fórmula de la *unión*,

$$P(A \cup E) = P(A) + P(E) - P(A \cap E) =$$
$$= 0{,}5 + 0{,}2 - 0{,}15 = 0{,}55$$

    b. Para que dos sucesos, A y E, sean independientes se debe cumplir que:

$$P(A \cap E) = P(A) \cdot P(E)$$

    En este caso, $P(A) \cdot P(E) = 0{,}5 \cdot 0{,}2 = 0{,}1$. Vemos que esta probabilidad es diferente de $P(A \cap E) = 0{,}15$. Por tanto, podemos concluir que NO son independientes.

    c. Nos preguntan ahora por la probabilidad $P(E / A)$. A partir de la fórmula de la probabilidad condicionada tenemos:

$$P(E / A) = \frac{P(E \cap A)}{P(A)} = \frac{0{,}15}{0{,}5} = 0{,}3$$

    Así, si nos informan que está lloviendo en Alemania, debemos “actualizar” la probabilidad que llueva en España, desde 0,2 hasta 0,3.

30. Una clase contiene igual número de estudiantes masculinos y femeninos. Después de un examen se observa que el 50% de los estudiantes ha obtenido un aprobado. Al mismo tiempo, el 30% de los exámenes obtienen un aprobado y han sido escritos por un varón. Averiguar la probabilidad que un examen con nota de aprobado haya sido contestado por un varón.

Este es otro ejemplo en que es apropiado usar la regla de Bayes. El conjunto de estudiantes puede dividirse en dos subconjuntos que son excluyentes y exhaustivos, E_1 el de los chicos y E_2 el de las chicas. Además al haber el mismo número de miembros en ambos subconjuntos tenemos que $P(E_1) = P(E_2) = 0{,}5$. También sabemos que la probabilidad de obtener un aprobado es del 50%, o sea, $P(A) = 0{,}5$ y que un 30% de los exámenes ha obtenido un aprobado y ha sido escrito por un chico, es decir, $P(A \cap E_1)=0{,}3$. La probabilidad buscada corresponde a que el examen haya sido contestado por un chico si sabemos que la nota es de aprobado:

$$P(E_1 / A) = \frac{P(E_1) \cdot P(A / E_1)}{P(A)} = \frac{P(E_1 \cap A)}{P(A)} = \frac{0{,}3}{0{,}5} = 60\%$$

31. Las empresas de seguros compilan estadísticas de accidentes de tráfico. Se sabe que aproximadamente el 10% de los conductores tienen más de un accidente de tráfico al año. Las compañías instalan en sus oficinas unos simuladores de conducción para que todos los conductores pasen un test de conducción (en el momento de la renovación de la póliza, por ejemplo). Después del primer año de funcionamiento del test, se comprueba que el 70% de los conductores que han tenido más de un accidente al año fallaron el test mientras que solo lo fallaron un 20% de los otros conductores. Con esta información las compañías de seguros quieren evaluar si aceptarán o no nuevos clientes, por lo que cualquier solicitante de una póliza debe realizar el test. Informar a la compañía de seguros para la que usted trabaja de cuál es la probabilidad que un solicitante que falle el test vaya a tener más de un accidente al año.

Vamos a distinguir dos tipos de conductores, los que tienen más de un accidente al año (E_1, los malos conductores) y los que como mucho solo tienen uno (E_2, los buenos). Estos dos tipos constituyen una clase de sucesos exhaustivos y excluyentes. Usaremos B para identificar el suceso “fallar el test”. De la información del texto sabemos:

$$P(E_1) = 0{,}1$$
$$P(E_2) = 0{,}9$$
$$P(B / E_1) = 0{,}7$$
$$P(B / E_2) = 0{,}2$$

Queremos evaluar la probabilidad que un conductor sea del tipo "malo" si sabemos que ha fallado el test. Entonces, usando la regla de Bayes hallaríamos:

$$P(E_1 / B) = \frac{P(E_1) \cdot P(B / E_1)}{P(E_1) \cdot P(B / E_1) + P(E_2) \cdot P(B / E_2)} = \frac{(0,1) \cdot (0,7)}{(0,1) \cdot (0,7) + (0,9) \cdot (0,2)} = 0,28$$

y observamos que la probabilidad *a priori* de 0,1 aumenta hasta 0,28 cuando se dispone de la información adicional de *B*.

32. (*) Tenemos tres dados indistinguibles externamente, pero uno de ellos está trucado con un peso de plomo que hace que la cara con el 6 aparezca la mitad de las veces. Se escoge un dado al azar, se lanza y se anota el número que aparece.

a. Si aparece un 6, decir cuál es la probabilidad que el dado sea el trucado.
b. Si aparece un 2, decir cuál es la probabilidad que el dado no sea el trucado.

a. Vamos a usar *T* para identificar el suceso "dado trucado" y *H* el suceso "dado honesto". Sabemos $P(T)=1/3$ y $P(H)=2/3$. Denominaremos *A* al suceso "aparece un 6 en el lanzamiento de un dado". Para el dado trucado sabemos que la probabilidad de observar un 6 será $P(A/T)=0,5$, mientras que si el dado es uno de los honestos tendremos $P(A/H)=1/6$. La probabilidad deseada es $P(T/A)$, que se puede evaluar usando la regla de Bayes:

$$P(T / A) = \frac{P(T) \cdot P(A / T)}{P(T) \cdot P(A / T) + P(H) \cdot P(A / H)} = \frac{(1/3) \cdot (1/2)}{(1/3) \cdot (1/2) + (2/3) \cdot (1/6)} = 0,60$$

b. En este caso haremos que *B* represente el suceso "aparece un 2", siendo la probabilidad buscada $P(H/B)$. Usando de nuevo la regla de Bayes:

$$P(H / B) = \frac{P(H) \cdot P(B / H)}{P(H) \cdot P(B / H) + P(T) \cdot P(B / T)} = \frac{(2/3) \cdot (1/6)}{(2/3) \cdot (1/6) + (1/3) \cdot (1/10)} = 0,77$$

El único aspecto de este cálculo que merece ser comentado es la probabilidad de que aparezca un 2 si se lanza el dado trucado. Para este dado, la probabilidad de obtener un 6 es de 0,5. El resto de caras acumularán también 0,5 de probabilidad, pero como hay 5 números distintos al 6, bajo el supuesto de que el único efecto perverso del dado se da en el número 6, la probabilidad debe repartirse entre los 5 resultados distintos. Así:

$$P(B / T) = \frac{0,5}{5} = \frac{1}{10}$$

33. (*) Tenemos una urna vacía y la vamos a llenar de bolas de colores en función del lanzamiento de una moneda honesta dos veces. Si en el primer lanzamiento sale cara (H) añadimos 2 bolas blancas (B) y si sale cruz (T) añadimos 4 bolas negras. Si en el segundo lanzamiento de la moneda sale cara (H) añadimos 3 bolas negras (N) y si sale cruz (T) 6 blancas (B). Después de los dos lanzamientos de la moneda se extrae una bola de la urna al azar.

a. Evaluar la probabilidad $P(N)$ que la bola extraída sea negra (N).
b. Si la bola extraída es negra (N), ¿cuál es la probabilidad que hayan salido dos caras (HH) en los lanzamientos de la moneda?

a. La bola extraída puede ser negra a través de cuatro caminos distintos: HH, HT, TH, TT. Según la regla estipulada para rellenar la urna tendríamos la siguiente composición de bolas de color negro (N) en cada una de esas cuatros rutas:

   HH: 0+3=3 (con 2 bolas blancas)
   HT: 0+0=0 (con 8 bolas blancas)
   TH: 4+3=7 (con 0 blancas)
   TT: 4+0=4 (con 6 blancas)

   La bola puede ser de color negro, en principio, siguiendo cualquiera de estas cuatro rutas. Usando la regla de la probabilidad total podemos calcular la probabilidad buscada:

$$P(N) = P(N/HH)\cdot P(HH) + P(N/HT)\cdot P(HT)$$
$$+ P(N/TH)\cdot P(TH) + P(N/TT)\cdot P(TT)$$

   Como la moneda es honesta sabemos que $P(HH)=P(HT)=P(TH)=P(TT)=0{,}25$. Sustituimos y hallamos:

$$P(N) = (3/5)\ 0{,}25 + (0/8)\ 0{,}25 + (7/7)\ 0{,}25 + (4/10)\ 0{,}25 = 0{,}50$$

b. Una vez observamos que la bola extraída es de color negro (N) podemos evaluar la probabilidad que ello sea así a través de la ruta HH usando la regla de Bayes:

$$P(HH/N) = \frac{P(N/HH)\cdot P(HH)}{P(N)} = \frac{(3/5)\cdot 0{,}25}{0{,}50} = 0{,}30$$

34. (*) El conductor típico de un automóvil tiene una probabilidad de sufrir un accidente del 1%. Esta probabilidad aumenta al 5% cuando llueve. Se sabe, además, que la probabilidad de lluvia en un día dado es del 24%.

a. Si nuestro conductor típico ha tenido un accidente de tránsito, ¿cuál es la probabilidad de que haya estado lloviendo?
b. ¿Cuál es la probabilidad de que no llueva o no tenga un accidente? (una pista: si R es el suceso "llueve" y A el suceso "tener un accidente", lo que buscamos es la probabilidad del suceso $\overline{R} \cup \overline{A}$.

a. Sabemos lo siguiente: $P(R)$=0,24. Sabemos también que la probabilidad de accidente aumenta cuando el tiempo pasa a ser lluvioso. Esto quiere decir que conocemos cómo aumenta la probabilidad de tener un accidente si hay lluvia:

$P(A/R)$ = 0,05 (cuando hay lluvia)
$P(A/\overline{R})$ = 0,01 (cuando no hay lluvia)

Deseamos evaluar la probabilidad $P(R/A)$ y lo podemos hacer usando la regla de Bayes:

$$P(R/A) = \frac{P(A/R) \cdot P(R)}{P(A)}$$

En esta expresión conocemos las probabilidades del numerador y solo necesitaríamos calcular la del denominador, y esta última la podemos evaluar usando la regla de la probabilidad total:

$$P(A) = P(A/R) \cdot P(R) + P(A/\overline{R}) \cdot P(\overline{R}) =$$
$$= (0,05) \cdot (0,24) + (0,01) \cdot (0,76) = 0,0196$$

Sustituyendo hallaríamos que $P(R/A)$ = 0,6122.

b. El suceso $\overline{R} \cup \overline{A}$ equivale al suceso $\overline{R \cap A}$ (se sigue de la ley de De Morgan). Usando esta relación, la regla del complementario y la regla multiplicativa tendríamos:

$$P(\overline{R} \cup \overline{A}) = P(\overline{R \cap A}) = 1 - P(R \cap A) = 1 - P(R) \cdot P(A/R)$$
$$= 1 - (0,24) \cdot (0,05) = 0,988$$

Probabilidad: combinatoria

35. Un banco tiene capacidad para sentar cuatro personas pero hay diez personas que quieren sentarse. Calcular de cuántas maneras distintas pueden sentarse esas diez personas.

Este ejercicio puede interpretarse de dos maneras diferentes. La primera es considerar que cuatro personas en concreto sentadas en órdenes diferentes constituyen, a efectos del recuento que se pide, cuatro maneras distintas de sentarse. Se trataría en este caso de calcular las permutaciones de 10 elementos en grupos de 4. La segunda posibilidad es considerar, por el contrario, que las mismas cuatro personas sentadas en órdenes diferentes constituyen, de hecho, una misma manera de sentarse. Se trata ahora de calcular las diferentes combinaciones de 10 elementos en grupos de 4.

Primer caso: permutaciones de 10 elementos en "listas ordenadas" de 4

$$P(10,4) = 10\cdot 9\cdot 8\cdot 7 = 5.040$$

Segundo caso: combinaciones de 10 elementos de "grupos" de 4

$$C(10,4) = \frac{10!}{4!\cdot(6!)} = \frac{10\cdot 9\cdot 8\cdot 7}{4\cdot 3\cdot 2\cdot 1} = 210$$

36. Considere los diez dígitos 0, 1, 2, 3, ..., 9. Calcular el número de números de cuatro cifras que pueden formarse con esos dígitos si:

a. los dígitos pueden repetirse.
b. los dígitos no pueden repetirse.
c. la última cifra es cero y los dígitos no pueden repetirse.

a. Podemos pensar que hay 4 posiciones (cada una ocupable por un dígito). Si los dígitos pueden repetirse, en la primera posición hay 10 dígitos posibles, en la segunda también, etc. Dado que hay 4 posiciones ellos implica que el número de números de 4 cifras vendrá dado por: $10\cdot 10\cdot 10\cdot 10=10^4$.

b. Si los dígitos no pueden repetirse, una vez uno de los 10 dígitos ocupa la primera posición, quedan 9 disponibles para la segunda, 8 para la tercera, y 7 para la cuarta. Se trata, por tanto, de calcular las diferentes *permutaciones* de 10 dígitos en "listas" (números) de 4 cifras:

$$\mathrm{P}(10,4) = 10\cdot 9\cdot 8\cdot 7 = 5.040$$

c. Finalmente, si la última cifra debe ser el cero y los dígitos no pueden repetirse, tenemos 9 dígitos (ya que el 0 no lo podemos utilizar) a colocar en 3 posiciones (ya que la última está ocupada por el 0). Por tanto, tenemos *permutaciones* de 9 dígitos en “listas” de 3 dígitos.

$$P(9,3) = 9 \cdot 8 \cdot 7 = 504$$

37. Una bibliotecaria tiene la tarea de poner unos libros en una estantería. Tiene cuatro libros de matemáticas, seis de física y dos de economía. Calcular de cuántas maneras distintas pueden ordenarse los libros si

   a. los libros de cada materia han de estar juntos.
   b. solamente los libros de matemáticas han de estar juntos.

   a. Si los libros de cada materia han de estar juntos, podemos ver el problema de la siguiente manera. Hay 3 “zonas”, una para cada materia, en la que los libros ocuparan posiciones. Por ejemplo, en la “zona” de matemáticas hay 4 libros que pueden ordenarse de 4! maneras, en la “zona” de física habrán 6! posibilidades y en la de economía 2! maneras. Ahora bien, cada una de las ordenaciones en la zona de matemáticas podrá agruparse con cada una de las de la zona de física y con cada una de las de la zona de economía. En consecuencia habrán:

$$(4!) \cdot (6!) \cdot (2!)$$

maneras distintas. Sin embargo, esto no es todo. El cálculo anterior presupone que la primera zona es la de matemáticas, la segunda la de física, y la tercera la de economía. Evidentemente no tiene por qué ser así. La primera zona puede reservarse a física, por ejemplo. Como hay 3 materias diferentes, hay 3! maneras de ordenar esas materias, y en cada una de ellas tendremos el número anterior de posibles ordenaciones de los libros. Al final, habrá un número:

$$[(4!) \cdot (6!) \cdot (2!)] \cdot 3!$$

de ordenaciones de libros distintas.

b. El trabajo de la bibliotecaria no ha acabado. Si únicamente los libros de matemáticas han de estar juntos, podemos ver esos 4 libros como una "unidad" de forma que hay 1 + 6 + 2 = 9 objetos distintos. Pueden ordenarse de 9! maneras. En cada una de estas maneras los libros de matemáticas están juntos aunque ellos también pueden ordenarse manteniendo la unidad de agrupación. Como hay 4, esa unidad se mantiene para cualquiera de las 4! maneras de ordenarlos. En definitiva, habrán:

$$(9!)\cdot(4!)$$

ordenaciones de libros bajo el prerrequisito que los de matemáticas estén juntos.

38. Queremos sentar cinco chicos y cuatro chicas en fila de forma que las chicas ocupen las posiciones pares. Calcular de cuántas maneras se pueden sentar.

Las cuatro chicas deberán ocupar las posiciones 2, 4, 6 y 8. Los cinco chicos deberán, por tanto, ocupar las posiciones 1, 3, 5, 7 y 9. Así, por un lado tenemos que calcular las diferentes permutaciones de 4 chicas en 4 posiciones (las pares) y por otro las diferentes permutaciones de 5 chicos en 5 posiciones (las impares). Está claro que estas permutaciones serán sin reposición, ya que una misma persona no puede ocupar dos posiciones.

$$P(4,4) = 4\cdot 3\cdot 2\cdot 1 = 24$$
$$P(5,5) = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120$$

Finalmente, debemos observar que para cada una de las 24 posibles permutaciones de las chicas tenemos 120 posibles permutaciones de los chicos. Por tanto, para hallar el total de permutaciones que se nos piden debemos multiplicar ambos resultados.

Total de permutaciones de chicas en posiciones pares y chicos en posiciones impares:

$$24\cdot 120=2.880$$

39. (*) Un estudiante de económicas tiene cinco monedas: una de 1 céntimo, una de 5 céntimos, una de 10 céntimos, una de 20 céntimos y una de 50 céntimos. ¿Cuántas sumas diferentes de dinero se pueden hacer con esas cinco monedas?

Si el estudiante elige una sola de las 5 monedas podrá hacer 5 sumas (triviales, claro, en este caso), pues puede elegir 1 objeto entre 5, es decir, $C(5,1)=5$ combinaciones. Si elige 2 de las 5 monedas, y como para la suma de los valores nominales el orden de aparición de las monedas es irrelevante, tendrá $C(5,2) = 10$ posibles sumas. En general, y siguiendo esta lógica, las sumas posibles serán:

$$\sum_{j=1}^{5} C(5, j) = 31$$

40. (*) Un departamento universitario tiene doce profesores, tres de los cuales son catedráticos y el resto son profesores titulares. Se ha de constituir un tribunal de cinco miembros para juzgar una tesis doctoral. Calcular cuántos tribunales pueden formarse si:

a. la composición del tribunal es libre.
b. ha de haber siempre un catedrático, exactamente.
c. han de haber siempre dos catedráticos, exactamente.
d. al menos ha de haber un catedrático.
e. el catedrático más antiguo forma parte de cualquier tribunal.
f. el catedrático más antiguo y el titular más joven forman parte de cualquier tribunal.
g. suponer que el director de tesis es un catedrático y que el director no puede formar parte del tribunal. Repetir (a), (b), (c) y (d).

Así pues, tenemos 12 profesores en total, que hemos de combinar en “comités” (tribunales) de 5 miembros bajo los supuestos que se enumeran en el enunciado del ejercicio. Nótese en primer lugar, que en cada uno de estos supuestos el orden de los profesores no será importante. Efectivamente, una vez constituido un tribunal este será el mismo independientemente del orden con que se nombre a sus miembros. Por tanto, en todos los casos, se tratará de calcular las diferentes *combinaciones* de profesores (catedráticos y titulares) que podrán formar los tribunales.

a. En este caso los 12 profesores son candidatos a ocupar cada una de las 5 plazas del tribunal. Por tanto, tenemos:

$$C(12,5) = \frac{12!}{5!7!} = \frac{12 \cdot 11 \cdot 10 \cdots 3 \cdot 2 \cdot 1}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 7 \cdot 6 \cdots 2 \cdot 1} = 792$$

b. En este caso, como la comisión ha de contener necesariamente un catedrático, este se debe escoger entre los tres disponibles y esto se puede hacer de C(3,1) maneras. A su vez, para las restantes 4 posiciones del tribunal tenemos 9 candidatos titulares a elegir 4 entre ellos. Esto se puede hacer de C(9,4) maneras. Finalmente la comisión al completo puede verse como las parejas formadas con los elegibles del primer grupo (el de los catedráticos) con los elegibles del segundo (los titulares). Por tanto, tenemos:

$$C(3,1)\cdot C(9,4)=\frac{3!}{1!2!}\cdot\frac{9!}{4!5!}=3\cdot 126=378$$

c. En este caso, siguiendo la misma lógica que en el apartado anterior, tendríamos:

$$C(3,2)\bullet C(9,3)=\frac{3!}{2!1!}\bullet\frac{9!}{3!6!}=3\bullet 84=252$$

d. Aquí la manera más eficiente de proceder es ver cuántos tribunales no tendrían ningún catedrático. Una vez obtenido este número se resta del total de comisiones y tendremos las comisiones en las que por lo menos hay un catedrático. Tendremos:

$$C(12,5)-C(9,5)=792-\frac{9!}{5!\bullet 4!}=792-126=666$$

e. En este caso, una vez el catedrático más antiguo ocupa un puesto en el tribunal, solo 11 profesores son candidatos a ocupar cada una de las 4 plazas libres del tribunal. Por tanto, tenemos:

$$C(11,4)=\frac{11!}{4!\bullet 7!}=\frac{11\bullet 10\bullet\bullet\bullet 3\bullet 2\bullet 1}{(4\bullet 3\bullet 2\bullet 1)\bullet(7\bullet 6\bullet\bullet\bullet 2\bullet 1)}=330$$

f. En este caso, una vez el catedrático más antiguo y el titular más joven ocupan un puesto cada uno en el tribunal, solo 10 profesores son candidatos a ocupar cada una de las 3 plazas libres del tribunal. Por tanto, tenemos:

$$C(10,3)=\frac{11!}{3!\bullet 7!}=\frac{10\bullet 9\bullet\bullet\bullet 3\bullet 2\bullet 1}{(3\bullet 2\bullet 1)\bullet(7\bullet 6\bullet\bullet\bullet 2\bullet 1)}=120$$

g. Este apartado se resuelve siguiendo exactamente las pautas anteriores pero reduciendo el número de catedráticos elegibles de 3 a 2.

41. (*) Tarde o temprano volveremos a tener elecciones presidenciales en Can Barça. Suponer que en ese momento aparecen 6 candidatos "de cambio" y 2 de "continuistas". La cadena de radio TOPEGUAY está pensando en realizar tertulias con los candidatos a raíz de la campaña.

a. ¿Cuántas tertulias diferentes se pueden hacer con 4 de los candidatos?
b. ¿Y cuántas de 4 candidatos si 2 son "de cambio" y 2 son "continuistas"?
c. ¿Cuál es la probabilidad que una tertulia de 4 candidatos elegida al azar tenga 3 candidatos "de cambio" y 1 "continuista?

a. Las tertulias son agrupaciones en las que el orden es irrelevante. Se corresponden a la idea de "comité", en otras palabras, a combinaciones. En este caso tendremos $C(8,4) = 70$ tertulias diferentes al disponer de 8 candidatos y 4 plazas en la tertulia.

b. Entre los 6 tertulianos "de cambio" se pueden escoger 2 de ellos según $C(6,2)$ pues de nuevo son agrupaciones en las que el orden no importa. Por su parte, y siguiendo los mismos principios, entre los 2 tertulianos "continuistas" se pueden escoger $C(2,2)$ agrupaciones diferentes. Entre la selección "de cambio" y la de "continuistas" se pueden establecer, en conjunto, $C(6,2)\cdot C(2,2)=15\cdot 1=15$ tertulias diferentes.

c. Para la nueva configuración usamos el mismo procedimiento que en b), pero cambiando la composición y retomamos la información obtenida en a) tendremos probabilidad:

$$p = \frac{C(6,3)\cdot C(2,1)}{C(8,4)} = \frac{20\cdot 2}{70} = 0,571$$

42. En el sorteo de los cuartos de final de la Champions League se han de aparejar 8 equipos. Suponga que entre estos equipos están el Barça y el Real Madrid. Evaluar la probabilidad de que estos dos equipos queden emparejados en cuartos de final si el proceso de selección es totalmente aleatorio, es decir, si se realiza una primera extracción de dos equipos y estos son el Barça y el Madrid.

Entre 8 equipos se pueden formar parejas de 2 (y el orden no importa, son "comités") en número igual a $C(8,2)=8!/(6!\cdot 2!) = 28$. Entre estos 28 "comités" solo hay 1 que corresponde a un emparejamiento Barça-Real Madrid. En consecuencia, la probabilidad de que se emparejen en primera instancia es 1/28=3,57%. Que acabe el Barça emparejado con el Real Madrid tras efectuar todo el sorteo tiene una probabilidad de 1/7 (1 caso sobre 7 casos posibles en total).

43. Consideremos un experimento de azar en el que todos los objetos que intervienen son "honestos". Primero se lanza un dado y se observa el número de puntos obtenidos (1, 2, 3, 4, 5 o 6). A continuación se lanzan un número de monedas igual al número de puntos obtenidos en el lanzamiento del dado.

a. Calcular la probabilidad que se observe una secuencia que contenga 4 caras.
b. Si sabemos que se han observado 4 caras, calcular la probabilidad que el número de puntos haya sido de 5.

a. Usaremos la siguiente notación: $P(4c/1)$ representará, por ejemplo, la probabilidad de observar 4 caras si en el lanzamiento del dado sale un 1. En este caso es evidente que $P(4c/1)=0$ pues el suceso "observar 4 caras" es imposible cuando solo se lanza una moneda. De manera similar podemos concluir que $P(4c/2)=P(4c/3)=0$. El suceso "$4c/4$" ya no es imposible y se observará con probabilidad:

$$P(4c/4)= (0{,}5)\cdot(0{,}5)\cdot(0{,}5)\cdot(0{,}5)=0{,}0625$$

El suceso "$4c/5$" se observará en situaciones distintas pues hay varias secuencias con 4 caras en el lanzamiento de 5 monedas. En concreto hay $C(5,4)$ secuencias de cuatro caras, cada una de ellas con probabilidad $(0{,}5)^4\cdot(0{,}5)$. En consecuencia, la probabilidad (binomial, en este caso) estará dada por:

$$P(4c/5)=\mathrm{C}(5,4)\cdot(0{,}5)^4\cdot(0{,}5)^1=0{,}1563$$

Una consideración similar nos permite evaluar la probabilidad de 4 caras en 6 lanzamientos:

$$P(4c/6)=\mathrm{C}(6,4)\cdot(0{,}5)^4\cdot(0{,}5)^2=0{,}2344$$

Finalmente estamos en condiciones de evaluar la probabilidad de observar 4 caras, "$4c$". Necesitamos usar la regla de la probabilidad total:

$$P(4c)=\sum_{j=1}^{6}P(4c/j)\cdot P(j)=0\cdot\frac{1}{6}+0\cdot\frac{1}{6}+0\cdot\frac{1}{6}+(0{,}1563)\cdot\frac{1}{6}+(0{,}2344)\cdot\frac{1}{6}=0{,}0755$$

b. Aquí debe usarse la regla de Bayes. Queremos evaluar $P(5/4c)$ y para ello podemos proceder así

$$P(5/4c) = \frac{P(4c/5)\cdot P(5)}{P(4c)} = \frac{(0{,}1563)\cdot(1/6)}{0{,}0755} = 0{,}3450$$

44. ***Aterriza como Puedas***. Un Jumbo 747 tiene cuatro motores de vuelo. Las estadísticas de vuelos indican que en uno de cada cien vuelos un motor falla. Si fallan a la vez más de dos motores durante un vuelo, el avión corre un serio riesgo de accidente.

a. Calcular la probabilidad de que un avión no sufra un accidente causado por la falla de los motores. ¿Se queda usted tranquilo?
b. Recalcule la probabilidad anterior en el caso que i) el fallo de un motor ocurriera en uno de cada diez vuelos, y ii) en uno de cada doscientos vuelos.

a. Cada uno de los 4 motores puede fallar o funcionar, con probabilidades respectivas de p y 1-p. En nuestro caso sabemos que p = 0,01. Podemos asumir, además, que el funcionamiento de cada motor es independiente del del resto de motores. Hagamos que F_i represente el suceso “fallan i motores exactamente” con i = 1, 2, 3, 4. Representaremos por A el suceso “riesgo serio de accidente” de forma que $A = F_3 \cup F_4$. Tres de los cuatro motores pueden fallar a la vez de $C(4,3)$ = 4 maneras; a cada una de estas maneras le corresponde una probabilidad de $p^3\cdot(1-p)$. Así, por tomar un ejemplo, la probabilidad que fallen 3 de los 4 motores viene dada por $P(F_3) = C(4,3)\cdot p^3\cdot(1-p)$. El fallo de los 4 motores a la vez solo puede producirse de una única manera que tiene probabilidad $P(F_4) = p^4$. De aquí es inmediato ver que:

$$P(A) = P(F_3 \cup F_4) = P(F_3) + P(F_4) = C(4,3)\cdot p^3\cdot(1-p) + p^4$$

puesto que $F_3 \cup F_4 = \varnothing$. Sustituyendo hayamos que $P(A)$ = 0,00000397 y por tanto $P(A)$ = 0,99999603. Puede usted volar tranquilo.

b. Para finalizar, rehaga usted mismo los cálculos para las dos probabilidades alternativas especificadas en este apartado usando los procedimientos delineados en el subapartado *a*.

45. Una caja contiene 8 bolas rojas, 3 bolas blancas y 9 azules. Considere el experimento aleatorio de extraer tres bolas al azar. Usando exclusivamente combinatoria, calcular la probabilidad de que

   a. Las tres bolas sean rojas
   b. Las tres sean blancas
   c. Las tres sean azules
   d. Dos sean rojas y una sea blanca
   e. Al menos una sea blanca
   f. Haya una de cada color
   g. Se extraigan en el orden roja, blanca, azul.

   Tenemos un total de 8 + 3 + 9 = 20 bolas. Si se extraen tres bolas al azar de entre veinte, el número de extracciones posibles es de $C(20,3)$, ya que el orden no importa.

   a. Puesto que hay 8 bolas rojas, que las 3 extraídas sean de ese color supone que ello puede ocurrir de $C(8,3)$ maneras. La probabilidad que las 3 sean rojas (casos a favor sobre casos totales) se puede evaluar como:

$$P = \frac{C(8,3)}{C(20,3)} = \frac{56}{1140} = 4,91\%$$

   b. Las tres blancas se pueden extraer de C(3,3) = 1 manera. En este caso:

$$P = \frac{C(3,3)}{C(20,3)} = \frac{1}{1140} = 0,088\%$$

   c. Como hay 9 bolas azules, siguiendo el mismo protocolo que arriba, ahora tendremos:

$$P = \frac{C(9,3)}{C(20,3)} = \frac{84}{1140} = 7,37\%$$

   d. Las dos rojas se han de escoger de entre las ocho de ese color y ello puede hacerse de $C(8,2)$ maneras. La bola blanca debe ser una de entre las tres de ese color y hay en consecuencia $C(3,1)$ maneras. Para constituir el trío de bolas, con dos rojas y una blanca, tenemos que ver las parejas posibles de dos rojas y una blanca y este número se obtiene de $C(8,2) \cdot C(3,1)$. La probabilidad buscada será:

$$P = \frac{C(8,2) \cdot C(3,1)}{C(20,3)} = \frac{28 \cdot 3}{1140} = 7,37\%$$

e. Habrá al menos una bola blanca en estos tres escenarios excluyentes: 1 blanca, 2 blancas, 3 blancas. Ahora no interesa distinguir el color del resto de bolas, es como si fuesen “no blancas” y de estas hay 8 + 9 = 17. El primer escenario corresponde a 1 bola blanca y 2 no blancas. Las maneras posibles son:

$$C(3,1)\cdot C(17,2)$$

Cuando se trata de dos bolas blancas el cálculo requerido será ahora de:

$$C(3,2)\cdot C(17,1)$$

Finalmente, para tres de color blanco tendremos:

$$C(3,3)\cdot C(17,0)$$

Así, recordando que los sucesos son excluyentes, la probabilidad deseada se puede evaluar usando la regla aditiva de la probabilidad mediante:

$$P=\frac{C(3,1)\cdot C(17,2)+C(3,2)\cdot C(17,1)+C(3,3)\cdot C(17,9)}{C(20,3)}$$

f. Ha de haber una bola de cada color. En este caso la de color blanco se puede extraer de $C(3,1)$ maneras, la de color rojo de $C(8,1)$ maneras y la de color azul de $C(9,1)$ maneras. Los tríos se obtendrán multiplicando y la probabilidad se evalúa por:

$$P=\frac{C(3,1)\cdot C(8,1)\cdot C(9,1)}{C(20,3)}$$

g. En el apartado anterior tenemos una bola de cada color, pero sin que el orden de aparición de las mismas sea relevante. Ahora sin embargo queremos ver la probabilidad que aparezcan en el orden roja-blanca-azul. Con los tres colores hay 3! maneras de que aparezcan en la extracción (por ejemplo, r-b-a, r-a-b, etc.). De estas 3! maneras solo hay una que interese (r-b-a). Por tanto, la probabilidad anterior, que corresponde a cualquier orden de aparición, debe ahora minorarse para el único caso que interesa, que es uno entre 3!. La probabilidad será ahora:

$$P=\frac{1}{3!}\cdot\frac{C(3,1)\cdot C(8,1)\cdot C(9,1)}{C(20,3)}$$

46. Determinar la probabilidad que aparezcan 3 seises en 5 lanzamientos de un dado.

Este es sencillo. La aparición de un "seis" puede verse como un "éxito" (y su no-aparición como un "fracaso"). La probabilidad de éxito en un lanzamiento es de 1/6, siempre que el dado sea honesto. Hay cinco repeticiones del suceso que, siendo independientes, permiten analizar la cuestión usando la variable binomial bajo el supuesto de querer la probabilidad de tres éxitos en cinco repeticiones. Por tanto:

$$P(\text{"3 seises en 5 lanzamientos"}) = C(5,3)\cdot(1/6)^3\cdot(5/6)^2 = 0{,}032$$

47. (*) En una escuela de música hay 4 cajonistas, 6 guitarristas, 3 bajistas y 2 saxofonistas. La maestra tiene que elegir una banda de 4 alumnos para la representación de la escuela en una competición internacional. Si la maestra elige los 4 miembros del grupo al azar, cuál es la probabilidad que:

a. En el grupo haya un cajonista, un guitarrista, un bajista y un saxofonista
b. En el grupo haya solo guitarristas
c. En el grupo haya al menos 2 guitarristas

a. Calculamos en primer lugar el número de resultados diferentes que componen el espacio muestral, es decir, el número de bandas de 4 miembros diferentes que podemos formar con los 15 músicos que hay en la clase. Dado que el orden no importa tendremos:

$$n_\Omega = C(15,4) = \frac{15!}{4!11!} = 1.365$$

[Nótese que este resultado nos servirá para los tres apartados, ya que el espacio muestral es siempre el mismo]

Calculamos ahora el número de resultados diferentes que componen el suceso.

A="En el grupo haya un cajonista, un guitarrista, un bajista y un saxofonista"

$$n_A = C(4,1)\cdot C(6,1)\cdot C(3,1)\cdot C(2,1) =$$
$$= 4\cdot 6\cdot 3\cdot 2 = 144$$

Por tanto, la probabilidad que se pide es:

$$P(A) = \frac{114}{1365} = 0{,}084$$

b. Calculamos el número de resultados diferentes que componen el suceso:

B="En el grupo haya solo guitarristas"

$$n_B = C(6,4) = 15$$

Por tanto, la probabilidad que se pide es:

$$P(B) = \frac{15}{1365} = 0{,}011$$

c. Nos piden la probabilidad del suceso

C="En la banda hay al menos dos guitarristas"

Lo podemos calcular por dos métodos distintos:

Método directo: si definimos los sucesos:

C_2 = "En la banda hay exactamente 2 guitarristas"
C_3 = "En la banda hay exactamente 3 guitarristas"
C_4 = "En la banda hay exactamente 4 guitarristas"

Vemos que $C = C_2 \cup C_3 \cup C_4$. Por tanto, dado que estos tres sucesos no tienen intersección entre sí tendremos que:

$$n_C = n_{C_2} + n_{C_3} + n_{C_4}$$

Calculando tenemos:

$$n_{C_2} = C(6,2) \cdot C(9,2) = 15 \cdot 30 = 540$$
$$n_{C_3} = C(6,3) \cdot C(9,1) = 20 \cdot 9 = 180$$
$$n_{C4} = C(6,4) \cdot C(9,0) = 15 \cdot 1 = 15$$

Por tanto, n_C = 540 + 180 + 15 = 735. En definitiva,

$$P(C) = \frac{735}{1365} = 0,538$$

Método complementario: si definimos los sucesos:

C_0 = "En la banda hay exactamente 0 guitarristas"
C_1 = "En la banda hay exactamente 1 guitarrista"

Vemos que $\bar{C} = C_0 \cup C_1$. Es decir, "En la banda hay al menos 2 guitarristas" es el complementario de "En la banda hay menos de 2 guitarristas", que es equivalente a "En la banda hay 0 guitarristas o en la banda hay 1 guitarrista". Por tanto, dado que estos dos sucesos no tienen intersección entre sí tendremos que:

$$P(C) = 1 - P(\bar{C}) = 1 - P(C_0 \cup C_1) = 1 - [P(C_0) + P(C_1)]$$

Calculando tenemos:

$$n_{C_0} = C(6,0) \times C(9,4) = 1 \times 126 = 126$$
$$n_{C_1} = C(6,1) \times C(9,3) = 6 \times 84 = 504$$

Por tanto,

$$P(\bar{C}) = P(C_0) + P(C_1) = \frac{126}{1365} + \frac{504}{1365} = 0,462$$

Finalmente,

$$P(C) = 1 - P(\bar{C}) = 1 - 0,462 = 0,538$$

48. ***Crimen y castigo***. La policía de fronteras revisa los camiones a fin de descubrir la entrada ilegal de objetos. Llega un camión con 10 contenedores que contienen 100 conejos de peluche cada uno. Suponga que usted sabe (a través de un soplo de un confidente policial) que en uno de los cajones hay 10 conejos rellenos de diamantes de contrabando. La policía, que desconoce el hecho, realiza una revisión rutinaria, razón por la que solo abrirá al azar uno de los contenedores y solo verificará tres de los conejos de peluche, también escogidos al azar.

   a. Informe al comisario de fronteras sobre la probabilidad que un registro rutinario del camión descubra el contrabando.
   b. Se produce una llamada anónima informando que en un camión de los que cruzarán la frontera con conejos de peluche habrá un cajón con contrabando. El comisario no tiene suficiente personal para registrar concienzudamente todos los camiones, así que contempla dos opciones. La primera consiste en abrir dos cajones al azar y comprobar tres de los conejos en cada cajón. La segunda consiste en abrir solo un cajón pero comprobar cuatro conejos al azar. ¿Qué curso de acción le aconsejaría usted tomar al comisario de entre estos dos?

   a. Cuando se elige uno de los contenedores hay $C(100,3)$ maneras de extraer tres conejos de peluche del mismo. Se descubre el contrabando en el caso que al menos uno de los tres conejos esté relleno de diamantes. Ello puede ocurrir de tres formas diferentes. En la primera hay un único conejo con diamantes, en la segunda hay dos y en la tercera hay tres. Las maneras en que puede ocurrir el primer caso se pueden contar teniendo en cuenta que habría uno de entre diez posibles con diamantes mientras que quedarían dos conejos de entre los otros 90 que no ocultan diamantes. De este tipo habrán un total de $C(10,1)\cdot C(90,2)$ tríos. En el caso que aparezcan dos conejos con diamantes ocultos, el número de posibilidades de este tipo de tríos es de $C(10,2)\cdot C(90,1)$. Finalmente tríos con los tres conejos ocultando diamantes habrá un total de $C(10,3)\cdot C(90,0)$. El total de tríos con detección positiva es en consecuencia de:

$$C(10,1)\cdot C(90,2)+C(10,2)\cdot C(90,1)+C(10,3)\cdot C(90,0)=$$
$$=40.050+4.050+120=44.220$$

   Puesto que el total de extracciones posibles es de $C(100,3)$ = 161.700, la probabilidad de detectar el contrabando una vez se ha seleccionado un contenedor es de (casos a favor/casos posibles):

   44.220/161.700=27,35%

   Ahora bien, hay 10 contenedores a escoger y solo uno de ellos es examinado. En este caso la probabilidad de detectar el contrabando debe ponderarse por la probabilidad de escoger el contenedor en el que efectivamente hay contrabando, que es de 1/10. En definitiva, la probabilidad buscada es:

   (1/10)×(44.200/161.700) = 2,74%

b. Se trata de repetir el apartado anterior bajo dos escenarios distintos. Para seleccionar dos contenedores de entre diez hay $C(10,2) = 45$ maneras. Como hay un único contendor con contrabando habrán $C(1,1)\cdot C(9,1) = 9$ maneras de que aparezca el contenedor con contrabando en la selección. Así, la probabilidad que la selección de dos contenedores contenga contrabando es de 9/45. Ahora toca abrir cada contenedor, extraer tres conejos y ver si contienen contrabando. Esta probabilidad, para un contenedor dado, ya ha sido calculada en el apartado anterior y su valor es 27,35%. Por tanto, la probabilidad de detectar contrabando es ahora de:

$$(9/45)\cdot(0{,}2735) = 5{,}47\%$$

La segunda opción consiste en seleccionar un contendor y revisar cuatro conejos. Es lo mismo que en el apartado anterior excepto por el número de conejos revisados. En este caso tendremos un total de $C(100,4)$=3.921.225 extracciones y entre estas con al menos un conejo con diamantes tendremos:

$$\begin{aligned}
&C(10,1)\cdot C(90,3)+C(10,2)\cdot C(90,2)+C(10,3)\cdot C(90,1)\\
&+C(10,4)\cdot C(90,0)=\\
&=1.174.800+180.225+10.800+210=1.366.035
\end{aligned}$$

La probabilidad será en consecuencia de:

$$(1/10)\cdot(1.366.035/3.921.225) = 3{,}48\%$$

En términos de probabilidad es más conveniente para el comisario seleccionar dos contenedores y examinar tres conejos que seleccionar uno y examinar cuatro.

49. ***Expediente X***. En el universo hay 100 mil millones de galaxias detectables. La Vía Láctea es una galaxia del todo ordinaria dentro de ese conjunto de galaxias y se sabe que contiene 150 mil millones de estrellas. Entre estas se calcula que solo 6 mil millones son del mismo tipo que el Sol. Sin embargo, no todas las estrellas similares al Sol contienen sistemas planetarios en los que pueda desarrollarse la vida. Se calcula que en solo uno de cada diez de estos sistemas se puede desarrollar vida o ser habitable. La presencia de vida no implica que esta sea necesariamente inteligente. Suponga que solo en uno de cada diez planetas con vida, esta ha llegado al nivel necesario para denominarse inteligente. La inteligencia no conlleva necesariamente la presencia de tecnología avanzada (solo hemos de mirar a la raza humana para comprobarlo). Suponga que en solo uno de cada diez planetas con vida inteligente, esta ha llegado a desarrollar tecnología avanzada.

a. Calcule la probabilidad de vida en la galaxia y en el universo.
b. Calcule la probabilidad de vida inteligente en la galaxia y en el universo.
c. Calcule la probabilidad de vida inteligente tecnológica en la galaxia y en el universo.
d. Con las probabilidades anteriores, estime el número de planetas que pueden contener vida de los tres tipos mencionados.

a. Nos centraremos en evaluar la probabilidad en la Vía Láctea. Siendo una galaxia típica, la probabilidad de vida en el universo será la misma que en la galaxia, pues al evaluar la probabilidad en términos de casos a favor sobre casos totales tanto en el numerador como en el denominador del universo se aplicaría el mismo factor corrector (i.e. el número de galaxias). Con los datos suministrados la probabilidad de vida en la galaxia es (casos a favor sobre casos totales) de:

$$P(V)=\frac{600 \text{ millones de estrellas habitables}}{150.000 \text{ millones de estrellas en la galaxia}}=0{,}004$$

b. A continuación hemos de evaluar la probabilidad que haya vida y sea inteligente, que puede evaluarse usando la regla multiplicativa:

$$P(V\cap I)=P(V)\cdot P(I/V)=(0{,}004)\cdot(0{,}10)=0{,}0004$$

c. Siguiendo pasos similares, evaluamos la probabilidad que hay vida y sea inteligente y haya desarrollado tecnología:

$$P(V\cap I\cap T)=P(V\cap I)\cdot P(T/V\cap I)=(0{,}004)\cdot(0{,}10)=0{,}00004$$

d. Finalmente, para evaluar el número de planetas en el universo que pueden albergar vida con las características descritas, partimos de cuántos planetas de esas características pueden existir en una galaxia típica como la nuestra (600 millones, 60 millones y 6 millones, respectivamente) y multiplicar por el número de galaxias en el universo (i.e. 100.000 millones). Así, y aceptando los datos suministrados, en el universo existirían (100.000 millones de galaxias) X (600 millones de planetas habitables por galaxia) = 60 trillones (60.000.000.000.000.000.000) millones de planetas habitables en el universo!! Un ejercicio interesante es reducir las ratios usadas pasando de 1/10 a 1/100 o a 1/1.000. Repita el cálculo y determine el número de planetas con las características de disponer de vida tecnológica.

50. (*) Un equipo de baloncesto está formado por cinco jugadores titulares. Cada uno de ellos habrá nacido en un lunes o un martes, etc. Evaluar la probabilidad que exista como mínimo una coincidencia en el día de nacimiento de los cinco jugadores del equipo (ello quiere decir que dos o más jugadores compartan el lunes o el martes, etc., como día de nacimiento).

El procedimiento para resolver esta cuestión es absolutamente idéntico al usado en el problema de los cumpleaños, pero ahora las coincidencias son en el nombre del día de la semana y no en el número del día del mes. Tomemos pues $n = 7$ (los siete días de la semana) y $r = 5$ (los cinco jugadores). Cada jugador del equipo puede haber nacido en uno de los 7 días de la semana y como hay cinco jugadores, tendremos que el total de todas las posibles combinaciones de días de nacimiento estará dado por:

$$n \cdot n \cdot n \cdot n \cdot n = n^r$$

Este número nos da la lista de posibles días de nacimiento para los cinco jugadores. Por otra parte, en esta lista las no-coincidencias se pueden enumerar calculando las permutaciones de los $n = 7$ días tomadas en bloques de $r = 5$. Hemos de calcular permutaciones, pues si el primer jugador hubiese, por ejemplo, nacido un lunes, ese día ya no podría ser considerado para el segundo jugador, si lo que deseamos es contar las no-coincidencias. En consecuencia los casos de no-coincidencia serían:

$$P(n,r) = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-(r-1))$$

La probabilidad p de que no haya ninguna coincidencia se evalúa como el cociente de los casos de no-coincidencia sobre los casos totales:

$$p = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-(r-1))}{n \cdot n \cdot n \cdot n \cdot n = n^r} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}{7^5} = 0{,}15$$

Usando la regla del complementario la probabilidad del suceso “por lo menos una coincidencia en el día de nacimiento” será de 1-p = 0,85, una probabilidad verdaderamente alta.

Variables aleatorias discretas

51. Considere el experimento aleatorio repetible de lanzar un dado dos veces y para las siguientes funciones (variables aleatorias) definidas en el espacio muestral del experimento obtenga las distribuciones de probabilidad correspondientes.

a. $X: \Omega \rightarrow R$ donde X mide las ganancias de un apostador que obtiene como premio de un lanzamiento tantos euros como puntos aparecen en el primer dado pero ha de pagar tantos euros como puntos aparecen en el segundo dado.

b. $X: \Omega \rightarrow R$ donde X mide las ganancias de un apostador que recibe como premio en euros la cifra más alta de los dos dados pero ha de pagar la cifra más pequeña.

c. $X: \Omega \rightarrow R$ donde X mide las ganancias de un apostador que recibe como premio en euros la cifra más alta que aparezca en las dos caras.

a. La siguiente tabla muestra los posibles resultados del lanzamiento de dos dados (D1 y D2) junto con el valor de las ganancias (D1-D2):

D1 / D2	*1*	*2*	*3*	*4*	*5*	*6*
1	0	-1	-2	-3	-4	-5
2	1	0	-1	-2	-3	-4
3	2	1	0	-1	-2	-3
4	3	2	1	0	-1	-2
5	4	3	2	1	0	-1
6	5	4	3	2	1	0

Vemos que hay 36 resultados equiprobables del lanzamiento de los dos dados, y cada uno de estos lanzamientos corresponde a un valor de la variable aleatoria. Sin embargo, algunos valores de la variable se repiten en más de un resultado. Así, contando en cuántos resultados aparece cada uno de los valores de las variables tenemos la siguiente distribución de probabilidad:

x	$p(x)$
-5	$1/36$
-4	$2/36$
-3	$3/36$
-2	$4/36$
-1	$5/36$
0	$6/36$
1	$5/36$
2	$4/36$
3	$3/36$
4	$2/36$
5	$1/36$

b. La siguiente tabla muestra los posibles resultados del lanzamiento de dos dados (D1 y D2) junto con el valor de las ganancias que se pueden expresar de forma algebraica usando la expresión MAX{D1, D2} – MIN{D1,D2}. En este caso al valor más alto se le resta el valor menor y cuando ambas caras coinciden una es atribuida como mayor y la otra como menor, de forma que en este caso la ganancia es nula:

D1 / D2	*1*	*2*	*3*	*4*	*5*	*6*
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

De nuevo hay 36 resultados equiprobables del lanzamiento de los dos dados, y cada uno de estos lanzamientos corresponde a un valor de la variable aleatoria. Como antes, contando en cuántos resultados aparece cada uno de los valores de la variable tenemos la siguiente distribución de probabilidad:

x	$p(x)$
0	$6/36$
1	$10/36$
2	$8/36$
3	$6/36$
4	$4/36$
5	$2/36$

c. Finalmente, la siguiente tabla muestra los posibles resultados del lanzamiento de dos dados (D1 y D2), junto con el valor de las ganancias medidas en este caso simplemente por la expresión MAX{D1, D2}:

D1 / D2	*1*	*2*	*3*	*4*	*5*	*6*
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Vemos que, nuevamente, hay 36 resultados equiprobables del lanzamiento de los dos dados, y cada uno de estos lanzamientos corresponde a un valor de la variable aleatoria. Como antes, contando en cuántos resultados aparece cada uno de los valores de la variable tenemos la siguiente distribución de probabilidad:

x	$p(x)$
1	$1/36$
2	$3/36$
3	$5/36$
4	$7/36$
5	$9/36$
6	$11/36$

52. Tenemos un grupo de 50 estudiantes de Estadística, 10 de los cuales son chicos y 40 son chicas. El profesor piensa regalar 3 matrículas de honor al azar a los estudiantes. En el espacio muestral de este experimento defina una función (variable aleatoria) X: $\Omega \rightarrow R$ que mida el número de matrículas asignables a los chicos.

   a. Obtener la distribución de probabilidad de X.
   b. Repetir el ejercicio si X midiera las matrículas asignables a las chicas.
   c. Repetir los dos apartados anteriores si el número de matrículas a regalar fuese 4.

   a. Si X = "Número de matrículas asignables a los chicos" tenemos que $X \in \{0,1,2,3\}$

   Calculamos en primer lugar el número de las diferentes posibles asignaciones de las tres matrículas, considerando tanto chicos como chicas. Esto es, cuántas combinaciones de 3 alumnos se pueden formar con los 50 del grupo:

$$C(50,3) = \frac{50!}{3!47!} = 19.600$$

Para obtener ahora la distribución de probabilidad de X debemos obtener, para cada $X \in \{0,1,2,3\}$

$$P(X = x) = \frac{\text{De cuantas maneras se pueden asignar x matrículas entre 10 chicos y } (3 - x) \text{ entre 40 chicas}}{\text{De cuantas maneras se pueden asignar } x \text{ matrículas entre los 50 del grupo}} = \frac{C(10,x) \cdot C(40,3-x)}{C(50,3)}$$

Calculando obtenemos:

$$P(X=0)=\frac{C(10,0)\cdot C(40,3)}{C(50,3)}=0{,}504$$
$$P(X=1)=\frac{C(10,1)\cdot C(40,2)}{C(50,3)}=0{,}398$$
$$P(X=2)=\frac{C(10,2)\cdot C(40,1)}{C(50,3)}=0{,}092$$
$$P(X=3)=\frac{C(10,3)\cdot C(40,0)}{C(50,3)}=0{,}006$$

b. En el caso de contar las matrículas asignadas a las chicas tendríamos el resultado simétrico, esto es:

$$P(X=x)=\frac{\text{De cuantas maneras se pueden asignar x matrículas entre 40 chicas y } (3-x) \text{ entre 10 chicos}}{\text{De cuantas maneras se pueden asignar } x \text{ matrículas entre los 50 del grupo}}=\frac{C(40,x)\cdot C(10,3-x)}{C(50,3)}$$

Calculando obtenemos:

$$P(X=0)=\frac{C(40,0)\cdot C(10,3)}{C(50,3)}=0{,}006$$
$$P(X=1)=\frac{C(40,1)\cdot C(10,2)}{C(50,3)}=0{,}092$$
$$P(X=2)=\frac{C(40,2)\cdot C(10,1)}{C(50,3)}=0{,}398$$
$$P(X=3)=\frac{C(40,3)\cdot C(10,0)}{C(50,3)}=0{,}504$$

c. Finalmente, si el número de matrículas a asignar es 4 tendríamos ahora que $X\in\{0,1,2,3,4\}$.

Para el caso de los chicos tendremos:

$$P(X=0)=\frac{C(10,0)\cdot C(40,4)}{C(50,4)}=0,397$$
$$P(X=1)=\frac{C(10,1)\cdot C(40,3)}{C(50,4)}=0,429$$
$$P(X=2)=\frac{C(10,2)\cdot C(40,2)}{C(50,4)}=0,152$$
$$P(X=3)=\frac{C(10,3)\cdot C(40,1)}{C(50,4)}=0,021$$
$$P(X=4)=\frac{C(10,4)\cdot C(40,0)}{C(50,4)}=0,001$$

Y para el caso de las chicas observamos el resultado simétrico:

$$P(X=0)=\frac{C(10,0)\cdot C(40,4)}{C(50,4)}=0,001$$
$$P(X=1)=\frac{C(10,1)\cdot C(40,3)}{C(50,4)}=0,021$$
$$P(X=2)=\frac{C(10,2)\cdot C(40,2)}{C(50,4)}=0,152$$
$$P(X=3)=\frac{C(10,3)\cdot C(40,1)}{C(50,4)}=0,429$$
$$P(X=4)=\frac{C(10,4)\cdot C(40,0)}{C(50,4)}=0,397$$

53. En una caja tenemos 5 cartas (un 2, 3 y 4 de copas y un 5 y 6 de espadas). Se realiza el experimento aleatorio de extraer dos cartas (sin reemplazamiento).

   a. ¿Cuántos puntos contiene el espacio muestral del experimento?
   b. Considere la variable aleatoria X que mide la suma de los valores numéricos de las dos cartas. Obtener su distribución de probabilidad.
   c. Considere a continuación que X solo asigna una suma positiva si las dos caras son del mismo tipo. Si no son del mismo tipo se repite la extracción. Obtener su distribución de probabilidad.
   d. Repetir los apartados anteriores si la extracción se realiza con reemplazamiento.
   e. Obtener la función de probabilidad acumulada en los casos anteriores.

a. Dado que tenemos 5 cartas diferentes y se extraen 2 cartas (sin reemplazamiento) tenemos 5 posibles resultados para la primera carta y 4 para la segunda, un total de 20 posibilidades. La extracción, por ejemplo, que contiene el 2 de copas y el 5 de espadas (2**C**, 5**E**) es distinta de la extracción (5**E**, 2**C**). El orden es relevante y el resultado se puede calcular como una permutación:

$$P(5,2) = \frac{5!}{(5-2)!} = 5 \cdot 4 = 20$$

b. La tabla siguiente muestra todas las posibilidades y los correspondientes resultados de la variable aleatoria X= "suma de los valores numéricos de las dos cartas" (una ~~x~~ indica que ese resultado no es posible al no existir reemplazamiento).

C1 / C2	**2C**	**3C**	**4C**	**5E**	**6E**
2C	~~x~~	5	6	7	8
3C	5	~~x~~	7	8	9
4C	6	7	~~x~~	9	10
5E	7	8	9	~~x~~	11
6E	8	9	10	11	~~x~~

Dado que los 20 posibles resultados son equiprobables tenemos que la probabilidad de cada uno de ellos es de 1/20.

Vemos que la variable aleatoria puede tomar cualquier valor en el conjunto {5,6,7,8,9,10,11}. La probabilidad de cada uno de estos valores la hallaremos multiplicando 1/20 por el número de resultados en los que aparece el valor en cuestión. Así tenemos:

$$f_X(x) = \begin{array}{ll} 2/20 & x = 5 \\ 2/20 & x = 6 \\ 4/20 & x = 7 \\ 4/20 & x = 8 \\ 4/20 & x = 9 \\ 2/20 & x = 10 \\ 2/20 & x = 11 \end{array}$$

c. En este caso tendremos la tabla siguiente. De nuevo las casillas marcadas con ~~x~~ son resultados que no son posibles:

C1 / C2	2C	3C	4C	5E	6E
2C	~~x~~	5	6	~~x~~	~~x~~
3C	5	~~x~~	7	~~x~~	~~x~~
4C	6	7	~~x~~	~~x~~	~~x~~
5E	~~x~~	~~x~~	~~x~~	~~x~~	11
6E	~~x~~	~~x~~	~~x~~	11	~~x~~

Ahora solo tenemos 8 posibles resultados equiprobables. Por tanto, la probabilidad de cada uno de ellos será de 1/8.

Vemos que la variable puede tomar ahora cualquier valor en el conjunto {5,6,7,11}. La probabilidad de cada uno de estos valores la hallaremos de nuevo multiplicando 1/8 por el número de resultados en los que aparece el valor en cuestión. Así tenemos:

$$f_X(x) = \begin{cases} 2/8 & x = 5 \\ 2/8 & x = 6 \\ 2/8 & x = 7 \\ 2/8 & x = 11 \end{cases}$$

d. Si la extracción se realiza con reemplazamiento tendremos inicialmente la tabla de resultados siguiente:

C1 / C2	2C	3C	4C	5E	6E
2C	4	5	6	7	8
3C	5	6	7	8	9
4C	6	7	8	9	10
5E	7	8	9	10	11
6E	8	9	10	11	12

De este modo tenemos 25 posibles resultados equiprobables, siendo la probabilidad de cada uno de ellos de 1/25. Los valores posibles de esta variable aleatoria están en el conjunto {4,5,6,7,8,9,10,11,12}. Obtenemos la probabilidad de cada uno de estos valores igual que antes y tenemos:

$$f_X(x) = \begin{cases} 1/25 & x=4 \\ 2/25 & x=5 \\ 3/25 & x=6 \\ 4/25 & x=7 \\ 5/25 & x=8 \\ 4/25 & x=9 \\ 3/25 & x=10 \\ 2/25 & x=11 \\ 1/25 & x=12 \end{cases}$$

Finalmente, si solo suman las cartas del mismo tipo los resultados posibles son:

C1 / C2	**2C**	**3C**	**4C**	**5E**	**6E**
2C	4	5	6	~~x~~	~~x~~
3C	5	6	7	~~x~~	~~x~~
4C	6	7	8	~~x~~	~~x~~
5E	~~x~~	~~x~~	~~x~~	10	11
6E	~~x~~	~~x~~	~~x~~	11	12

Ahora tenemos 13 posibles resultados equiprobables. Por tanto, la probabilidad de cada uno de ellos será de 1/13. La variable puede tomar cualquier valor en el conjunto {4,5,6,7,8,10,11,12}. La probabilidad de cada uno de estos valores será ahora de:

$$f_X(x) = \begin{cases} 1/13 & x=4 \\ 2/13 & x=5 \\ 3/13 & x=6 \\ 2/13 & x=7 \\ 1/13 & x=8 \\ 1/13 & x=10 \\ 2/13 & x=11 \\ 1/13 & x=12 \end{cases}$$

e. Nos centraremos en la función de probabilidad acumulada para la primera de las variables aleatorias del apartado b.; para el resto se procedería exactamente igual y lo dejamos para el lector. Recordemos que la función de probabilidad acumulada está definida como:

$$F_X(x) = P(X \leq x)$$

De manera que tendremos:

$$F_X(x) = \begin{array}{cr} 2/20 & x=5 \\ 2/20 + 2/20 = 4/20 & x=6 \\ 2/20 + 2/20 + 4/20 = 8/20 & x=7 \\ 2/20 + 2/20 + 4/20 + 4/20 = 12/20 & x=8 \\ 2/20 + 2/20 + 4/20 + 4/20 + 4/20 = 16/20 & x=9 \\ 2/20 + 2/20 + 4/20 + 4/20 + 4/20 + 2/20 = 18/20 & x=10 \\ 2/20 + 2/20 + 4/20 + 4/20 + 4/20 + 2/20 + 2/20 = 20/20 = 1 & x=11 \end{array}$$

54. Considere una variable aleatoria discreta X que toma valores x=1, 2, 3 con idéntica probabilidad.
    a. Obtener la función de probabilidad $f_X(x)$ y la función cumulativa de probabilidad $F_X(x)$.
    b. Obtener la esperanza matemática $E(X)$ y la varianza var(X) de la variable aleatoria.
    c. Considerar la función $g: R \to R$ definida como g(x)=2x+3. Obtener la distribución de probabilidad de la variable aleatoria $g(X)$ así como su esperanza y varianza.

a. Al ser todos los valores equiprobables tendremos la función de probabilidad:

$$f_X(x) = \begin{cases} 1/3 & x = 1 \\ 1/3 & x = 2 \\ 1/3 & x = 3 \end{cases}$$

En cuanto a la función de distribución, dado que $F_X(x) = \sum_{t \le x} f_X(t)$ tenemos:

$$F_X(x) = \begin{matrix} 0 & x < 1 \\ 1/3 & 1 \le x < 2 \\ 2/3 & 2 \le x < 3 \\ 1 & x \ge 3 \end{matrix}$$

b. Para hallar la esperanza $E(X)$ debemos calcular:

$$E(X) = \sum_x x \cdot f_X(x)$$

En este caso,

$$E(X) = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 = 2$$

Para hallar la varianza var(X) podemos usar la fórmula:

$$\operatorname{var}(X) = \sum_x x^2 \cdot f_X(x) - E(X)^2$$

Y en este caso,

$$\operatorname{var}(X) = \frac{1}{3} \cdot 1^2 + \frac{1}{3} \cdot 2^2 + \frac{1}{3} \cdot 3^2 - 2^2 = \frac{2}{3}$$

c. Dado que $X \in \{1,2,3\}$ tendremos que $g(X) = 2X + 3 \in \{5,7,9\}$ con la siguiente asociación:

$$\begin{aligned} g(x) = 5 &\quad \Leftrightarrow \quad x = 1 \\ g(x) = 7 &\quad \Leftrightarrow \quad x = 2 \\ g(x) = 9 &\quad \Leftrightarrow \quad x = 3 \end{aligned}$$

Por tanto,

$$f_{g(X)}(g(x)) = \begin{array}{ll} 1/3 & g(x) = 5 \\ 1/3 & g(x) = 7 \\ 1/3 & g(x) = 9 \end{array}$$

En consecuencia,

$$E(g(X)) = \frac{1}{3} \cdot 5 + \frac{1}{3} \cdot 7 + \frac{1}{3} \cdot 9 = 7$$

Este resultado también se podría haber obtenido a partir de las propiedades de la esperanza matemática:

$$E(g(X)) = E(2 \cdot X + 3) = 2 \cdot E(X) + 3 = 2 \cdot 2 + 3 = 7$$

De la misma manera,

$$\text{var}(X) = \frac{1}{3}\cdot 5^2 + \frac{1}{3}\cdot 7^2 + \frac{1}{3}\cdot 9^2 - 7^2 = \frac{8}{3}$$

O bien,

$$\text{var}(g(X)) = \text{var}(2\cdot X + 3) = 4\cdot \text{var}(X) = 4\cdot\frac{2}{3} = \frac{8}{3}$$

55. Considere una variable aleatoria discreta *X* que mide la demanda diaria de un determinado producto. La distribución de probabilidad de *X* viene dada por:

$$P(X = x) = c\cdot\frac{2^x}{x!} \qquad x = 1,2,3,4$$

Suponer que el producto genera un ingreso de 5 euros por unidad vendida y que no hay otro coste que el de almacenar las unidades no vendidas. El fabricante produce diariamente *k* unidades (*k* = 1, 2, 3, 4) con un coste de almacenamiento de 3 euros por unidad no vendida.

a. Calcular el coeficiente *c*.
b. Obtener la demanda esperada.
c. Obtener la varianza de la demanda.
d. Obtener la distribución de probabilidad del beneficio diario en función del output.
e. Averiguar el nivel de producción que maximiza el beneficio diario esperado.

a. Sabemos que el total de probabilidad es necesariamente 1. En consecuencia debe cumplirse:

$$1 = \sum_x P(X = x) = \sum_x c\cdot\frac{2^x}{x!} = c\cdot\sum_x\frac{2^x}{x!}$$

Y de aquí se sigue que el valor de *c* que verifica la propiedad es:

$$c = \left(\sum_x \frac{2^x}{x!}\right)^{-1} = \frac{1}{6}$$

una vez hechos los cálculos pertinentes. Con esta constante la distribución de probabilidad es:

$$P(X=1)=\frac{1}{3}$$
$$P(X=2)=\frac{1}{3}$$
$$P(X=3)=\frac{8}{36}$$
$$P(X=4)=\frac{16}{144}$$

Sumando se verifica que efectivamente el total de probabilidad es 1.

b. Calculamos la esperanza matemática y hallamos:

$$E(X)=\sum_x x\cdot P(X=x)=2{,}111$$

c. A continuación calculamos la varianza para obtener:

$$\mathrm{var}(X)=E(X^2)-E(X)^2=\sum_x x^2\cdot P(X=x)-E(X)^2=0{,}988$$

d. Quizás la mejor manera de visualizar la situación es preparando una tabla en la que por filas aparecen los posibles valores de la demanda (una variable aleatoria) y por columnas las producciones (una variable de decisión de la empresa) y en las celdas tenemos el beneficio en función de la materialización de la demanda y la decisión de la empresa. Así en la celda $x = 2$, $k = 3$ aparece una cifra de beneficio 7 obtenida por la diferencia entre el ingreso neto, $2 \times 5 = 10$, de las dos unidades compradas por la demanda menos el coste de la unidad no vendida y almacenada, $(3 - 2) \times 3 = 3$. Y así sucesivamente para el resto de las celdas de la tabla. Obsérvese que si la empresa produce $k = 1$, entonces en todos los escenarios de demanda el beneficio es 5 pues como mucho los demandantes solo podrán adquirir la única unidad producida por la empresa.

x / k	***1***	***2***	***3***	***4***
1	5	2	-1	-4
2	5	10	7	4
3	5	10	15	12
4	5	10	15	20

Vamos a proceder ahora a estudiar la distribución de probabilidad del beneficio en función de la decisión de la empresa. Para k = 1, el beneficio es 5 con probabilidad 1. Para k = 2 el beneficio es 2 si la demanda es x = 1 y 10 en los casos contrarios, etc. Usando $B(k)$ como el beneficio según output k tendremos:

$$B(1) = 5 \quad \text{con probabilidad} = 1$$

$$B(2) = \begin{matrix} 2 & \text{con } P(X=1) = 1/3 \\ 10 & \text{con } P(X>1) = 2/3 \end{matrix}$$

$$B(3) = \begin{matrix} -1 & \text{con } P(X=1) = 1/3 \\ 7 & \text{con } P(X=2) = 1/3 \\ 15 & \text{con } P(X>2) = 1/3 \end{matrix}$$

$$B(4) = \begin{matrix} -4 & \text{con } P(X=1) = 1/3 \\ 4 & \text{con } P(X=2) = 1/3 \\ 12 & \text{con } P(X=3) = 8/36 \\ 20 & \text{con } P(X=4) = 16/144 \end{matrix}$$

e. Bajo las cuatro posibles decisiones de la empresa podemos evaluar el beneficio esperado pues disponemos de las probabilidades necesarias. Así, obtendremos:

$$E(B(1)) = 5$$
$$E(B(2)) = 7{,}333$$
$$E(B(3)) = 7$$
$$E(B(4)) = 4{,}888$$

Si la decisión de la empresa se basa en maximizar el beneficio esperado diario, la opción óptima seria producir k = 2.

56. (*) Considere dos experimentos aleatorios (o juegos de azar). El primero consiste en lanzar cuatro monedas honestas al aire. Introducimos una variable aleatoria X que otorga en euros el doble del número de caras que aparezcan en la realización del juego. El segundo consiste en lanzar una moneda honesta al aire y en este juego se introduce una variable aleatoria Y que otorga un premio de 8 euros en caso de cara y nada en caso contrario.

a. Si pudiera escoger en qué juego participar, ¿cuál elegiría basándose en la esperanza matemática de las variables aleatorias?

b. Suponga a continuación que el criterio de decisión fuese la utilidad de los premios obtenibles. ¿Qué juego elegiría si: i) $u(x) = \sqrt{x}$, ii) $u(x) = x$, iii) $u(x) = x^2$
c. En terminología económica, ¿qué tipo de actitud delante del riesgo tendríamos en cada uno de los tres casos?

a. La variable aleatoria X en cuanto mide el número de caras refleja un fenómeno binomial en el que $X = 0, 1, 2, 3, 4$ con probabilidad de éxito (salir cara) de 0,5. Las probabilidades se pueden calcular en consecuencia por:

$$P(X = x_i) = C(4, x_i) \cdot 0,5^{x_i} \cdot 0,5^{4-x_i}$$

con $x_i = 0$, etc. El premio en euros, que vamos a representar por G_X, está dado por $G_X = 2 \cdot X$ de forma que su esperanza matemática se calculará por:

$$E(G_X) = \sum_i g_i \cdot P(G_X = g_i) = \sum_i 2 \cdot x_i \cdot P(X = x_i)$$

que puede verificarse genera un valor esperado $E(G) = 4$. Para el segundo experimento tenemos una variable aleatoria binaria con $Y = 8$ con probabilidad 0,5 e $Y = 0$ con la misma probabilidad. Las ganancias en este caso tienen esperanza matemática:

$$E(G_Y) = 8 \cdot 0,5 + 0 \cdot 0,5 = 4$$

Desde la perspectiva de la esperanza matemática ambas situaciones de azar dan lugar a la misma esperanza matemática y tienen el mismo valor económico.

b. Pasemos a considerar ahora que lo relevante es la utilidad de los premios. Si la utilidad es lineal, es decir $u(x) = x$, estamos en la situación del apartado anterior. En el caso en que no es lineal, por ejemplo, si:

$$u(x) = \sqrt{x}$$

entonces hemos de calcular la utilidad esperada (o esperanza matemática de las utilidades de los premios):

$$E(u(G_X)) = \sum_i u(g_i) \cdot P(G_X = g_i) = \sum_i \sqrt{2 \cdot x_i} \cdot P(X = x_i)$$

Se puede verificar, sustituyendo los premios y las probabilidades requeridas, que en este caso la utilidad esperada del primer tipo de experimento asciende a:

$$E(u(G_X)) = 1,892$$

En el segundo experimento, es también fácil de verificar que la utilidad esperada asciende a:

$$E(u(G_Y)) = \sqrt{8} \cdot 0,5 + \sqrt{0} \cdot 0,5 = 1,414$$

Así pues, la primera situación de azar sería preferida a la segunda teniendo en cuenta que la utilidad esperada que reporta es mayor. Para la función de utilidad del tipo $u(x) = x^2$ dejamos que realice usted el cálculo.

c. Con utilidad lineal el comportamiento se cataloga de "neutral" delante del riesgo. Para la función $u(x) = \sqrt{x}$ tendríamos un comportamiento "adverso" al riesgo mientras que alguien con una utilidad como $u(x) = x^2$ sería descrito como "amante" del riesgo.

57. Un conductor desea vender su viejo coche, razón por la que considera gastar 50 euros poniendo un anuncio en una web especializada de compra/venta de coches usados. El precio de venta del coche es de 750 euros. Sin publicidad la probabilidad de vender el coche es del 50%, mientras que con publicidad el conductor sabe que nueve de cada diez coches puestos a la venta en la web son efectivamente vendidos. Si no consigue vender el coche, lo cederá a un amigo que le ofrece 650 euros. Discuta si el vendedor debe contratar el anuncio o no.

Hay dos situaciones a valorar y comparar: "contratar el anuncio", o bien "no contratar el anuncio". Si no se efectúa el anuncio, la esperanza para el vendedor es de:

$$\begin{aligned} E(\text{no contratar}) &= \text{vende al público} \cdot 0,5 + \text{vende al amigo} \cdot 0,5 = \\ &= 750 \cdot 0,5 + 650 \cdot 0,5 = 700 \end{aligned}$$

Si se contrata el anuncio aumenta la probabilidad de vender el coche al público, aunque entonces debe descontarse el coste del anuncio. En este caso:

$$\begin{aligned} E(\text{contratar}) &= \text{vende al público} \cdot 0,9 + \text{vende al amigo} \cdot 0,1 = \\ &= (750\text{-}50) \cdot 0,9 + (650\text{-}50) \cdot 0,1 = 690 \end{aligned}$$

Si al individuo en cuestión toma la esperanza matemática como criterio de decisión, le conviene no contratar el anuncio.

58. El profesor de Estadística va a poner un examen final tipo test. El examen va a consistir en diez preguntas de tipo “verdadero/falso” y de cinco preguntas de tipo “elección múltiple” en las que se ofrecen cuatro respuestas posibles. El profesor quiere que el examen refleje que un alumno que no ha estudiado absolutamente nada no pueda esperar otra nota que un cero cuando el alumno responda al azar. Todas las preguntas se han de contestar. Si cada pregunta acertada puntúa +1, calcule las puntuaciones de las preguntas no acertadas a fin de que se cumpla el criterio del profesor, si en el apartado de elección múltiple:

   a. Solo hay una respuesta correcta entre las cuatro.
   b. Hay dos respuestas correctas entre las cuatro.
   c. Puede haber más de una respuesta correcta entre las cuatro.

   a. Dado que las preguntas son de diferente tipo se deben puntuar de forma diferente.

   Preguntas de tipo “verdadero/falso”
   Definimos la variable X = “Puntuación de la pregunta”. Si denotamos con “r” la puntuación dada a una respuesta incorrecta tendremos:

$$X = \begin{cases} 1 & \text{si la respuesta es correcta} \\ r & \text{si la respuesta es incorrecta} \end{cases}$$

   Dado que la pregunta es del tipo “verdadero/falso” tenemos que una y solo una de las respuestas es correcta. Además suponemos que el estudiante contesta totalmente al azar. Por ello la probabilidad de obtener puntuaciones de 1 y de r es totalmente simétrica:

$$f_X(x) = \begin{cases} 1/2 & x = 1 \\ 1/2 & x = r \end{cases}$$

   Por tanto, si buscamos el valor esperado,

$$E(X) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot r = \frac{1+r}{2}$$

El profesor quiere que esta esperanza sea cero. Por tanto,

$$E(X)=\frac{1+r}{2}=0 \Rightarrow 1+r=0 \Rightarrow r=-1$$

Es decir, para conseguir que un estudiante que responde completamente al azar espere una nota de cero en las preguntas del tipo “verdadero/falso”, el profesor debe penalizar con 1 punto cada pregunta incorrecta (-1).

Preguntas de tipo “opción múltiple con 4 opciones”

Definimos la variable X = “Puntuación de la pregunta”. Si de nuevo denotamos con “r” la puntuación dada a una respuesta incorrecta tendremos:

$$X = \begin{matrix} 1 & \text{si la respuesta es correcta} \\ r & \text{si la respuesta es incorrecta} \end{matrix}$$

Dado que la pregunta es del tipo “opción múltiple con 4 opciones” y solo una de las 4 opciones es correcta (y estamos suponiendo que el estudiante contestará totalmente al azar), tendremos que la función de probabilidad sería en este caso:

$$f_X(x)=\begin{cases} 1/4 & x=1 \\ 3/4 & x=r \end{cases}$$

Por tanto, si buscamos el valor esperado:

$$E(X)=\frac{1}{4}\cdot 1+\frac{3}{4}\cdot r=\frac{1+3r}{4}$$

El profesor quiere que esta esperanza sea cero. Por tanto,

$$E(X)=\frac{1+3r}{4}=0 \Rightarrow 1+3r=0 \Rightarrow r=-\frac{1}{3}$$

Es decir, para conseguir que un estudiante que responde completamente al azar espere una nota de cero en las preguntas del tipo “opción múltiple”, el profesor debe penalizar con 1/3 de punto cada pregunta incorrecta (-1/3).

b. Si hay dos preguntas correctas entre las cuatro de opción múltiple

Preguntas de tipo "verdadero/falso"
Nada cambia con respecto al apartado a. anterior para las preguntas del tipo "verdadero/falso".

Preguntas de tipo "opción múltiple con 4 opciones"
Definimos de nuevo la variable X = "Puntuación de la pregunta". Si denotamos con "r" la puntuación dada a una respuesta incorrecta tendremos:

$$X = \begin{cases} 1 & \text{si la respuesta es correcta} \\ r & \text{si la respuesta es incorrecta} \end{cases}$$

Dado que la pregunta es del tipo "opción múltiple con 4 opciones" y ahora 2 de las 4 opciones son correctas (y estamos suponiendo que el estudiante contestará totalmente al azar), tendremos que:

$$f_X(x) = \begin{cases} 2/4 & x = 1 \\ 2/4 & x = r \end{cases}$$

Por tanto, si buscamos ahora el valor esperado,

$$E(X) = \frac{2}{4} \cdot 1 + \frac{2}{4} \cdot r = \frac{2 + 2r}{4} = \frac{1 + r}{2}$$

El profesor quiere que esta esperanza sea cero. Por tanto,

$$E(X) = \frac{1 + r}{2} = 0 \Rightarrow 1 + r = 0 \Rightarrow r = -1$$

Es decir, para conseguir que un estudiante que responde completamente al azar espere una nota de cero en las preguntas del tipo "opción múltiple", el profesor debe penalizar con 1 punto cada pregunta incorrecta (-1).

c. Si hay "más de una" repuestas correctas entre las cuatro de opción múltiple

Preguntas de tipo "verdadero/falso"
Nada cambia con respecto al apartado a. anterior para las preguntas del tipo "verdadero/falso".

Preguntas de tipo "opción múltiple con 4 opciones"
Definimos de nuevo la variable X = "Puntuación de la pregunta". Si denotamos una vez más con "*r*" la puntuación dada a una respuesta incorrecta tendremos:

$$X = \begin{cases} 1 & \text{si la respuesta es correcta} \\ r & \text{si la respuesta es incorrecta} \end{cases}$$

Ahora en las preguntas del tipo "opción múltiple con 4 opciones" se da el caso que "más de una" de las 4 opciones son correctas. Denotemos con *n* el número de opciones correctas. Bajo el supuesto adicional que el estudiante sabe el valor de *n* en cada una de las preguntas (por ejemplo, porque el profesor anuncia que hay 3 respuestas correctas, en este caso *n* = 3). Entonces tendremos que:

$$f_X(x) = \begin{cases} n/4 & x = 1 \\ 4 - n/4 & x = r \end{cases}$$

Por tanto, si buscamos el valor esperado:

$$E(X) = \frac{n}{4}\cdot 1 + \frac{4-n}{4}\cdot r = \frac{n + (4-n)\cdot r}{4}$$

Ahora bien, el profesor quiere que esta esperanza sea cero. Por tanto,

$$E(X) = \frac{n + (4-n)\cdot r}{2} = 0 \Rightarrow n + (4-n)\cdot r = 0 \Rightarrow r = -\frac{n}{4-n}$$

Es decir, para conseguir que un estudiante que responde completamente al azar espere una nota de cero en las preguntas del tipo "opción múltiple", el profesor debe penalizar con *n*/(4-*n*) punto cada pregunta incorrecta. Puede verse que cuando *n* = 2, obtenemos de nuevo el resultado del apartado previo en que *r* = -1.

Un escenario distinto es que el estudiante sabe que hay más de una respuesta correcta, pero el profesor no informa del número exacto de respuestas correctas. El estudiante, delante de cada pregunta, ha de indicar 2, 3 o 4 de las respuestas listadas. Para el estudiante, al carecer de información, y haber de responder al azar, lo que es relevante es de cuántas maneras puede señalar sus respuestas. Si señala 2 al azar va a tener *C*(4,2) = 6 maneras distintas de hacerlo. Si señala 3, entonces serán $C(4,3) = 4$ y si señala 4 entonces solo habrá una manera, $C(4,4) = 1$. En total hay 11 maneras de señalar respuestas al azar, y solo una de ellas será la correcta. La función de probabilidad será en este caso:

$$f_X(x) = \begin{cases} 1/11 & x = 1 \\ 10/11 & x = r \end{cases}$$

Buscamos el valor de r que haga que la esperanza de nota sea cero con esta distribución de probabilidad:

$$E(X) = 1 \cdot \frac{1}{11} + r \cdot \frac{10}{11} = 0 \quad \Rightarrow \quad r = -\frac{1}{10}$$

59. Una tabla de mortalidad establece que la probabilidad que un individuo de 30 años sobreviva durante el próximo año es del 0,992. Una compañía de seguros le ofrece la siguiente póliza de seguros: Usted paga una prima de 200 euros anuales a cambio de recibir (no usted, claro, sino sus herederos) 10.000 euros, en caso de muerte.

    a. Establezca si esta póliza es conveniente para la compañía.
    b. Establezca el valor de la prima que hace que la póliza sea “justa” (la empresa obtiene beneficio nulo).
    c. Repetir los apartados anteriores si cada póliza comporta unos costes de tramitación de 50 euros anuales.

    a. Vamos a llamar B al beneficio de la compañía de seguros. En términos de esperanza matemática, la compañía ha de tener en cuenta que la póliza le representa un ingreso de 200 euros si el contratante permanece en este mundo, mientras que ha de desembolsar 10.000 euros en caso contrario. Por tanto,

$$E(B) = 200 \cdot (0{,}992) + (-10.000) \cdot (0{,}008) = 118{,}4$$

En promedio, la compañía obtiene 118,4 euros por cada una de estas pólizas.

    b. Hemos definido la póliza “justa” como aquella que da lugar a un beneficio esperado de cero. Si q es dicho valor se debería verificar:

$$E(B) = q \cdot (0{,}992) + (-10.000) \cdot (0{,}008) = 0$$

de donde se deriva que con una prima de $q = 80{,}65$ euros la compañía no incurriría en pérdidas, en promedio.

c. Se deben replicar los cálculos previos cuando hay un coste de tramitación de c = 50 euros anuales. En el primer caso el beneficio si el contratante pervive sería de 200 – c, mientras que en caso contrario sería de –(10.000+c).

60. Considere un juego de azar en el que el jugador puede escoger un número del 1 al 6. A continuación se lanzan tres dados honestos. Si uno de los dados muestra el número elegido, el jugador recibe 100 euros, si dos de los dados muestran el número, recibe 200 euros y si los tres dados muestran el número, entonces recibe 300 euros. Si ningún dado aparece con el número elegido, el jugador pierde 100 euros. Tómese una variable aleatoria X que mida las veces que aparece el número preseleccionado.

a. Obtenga la *distribución* de probabilidad de X.
b. Obtenga la probabilidad que el jugador obtenga una ganancia positiva (es decir, $P(X > 0)$).
c. Obtenga la esperanza matemática del juego (o sea, de la variable aleatoria X).
d. Obtenga la varianza del juego (o sea, de X).
e. Suponga ahora que si ningún dado muestra el número elegido no se produce ninguna pérdida. Calcular el precio "justo" de un billete de lotería que ofreciera el derecho a participar en el juego. Si usted valora el dinero según la función de utilidad $u(x) = \sqrt{x}$, si $x > 0$, y $u(x) = -\sqrt{x}$, si $x < 0$, determine cuál es para usted el "valor" del juego.

a. Tenemos que X puede tomar valores X = 0, 1, 2, 3. Por ejemplo, X = 1 significa que el número elegido por el apostante aparece una vez en tres lanzamientos del dado. El apostante elige un número al azar entre 1 y 6. Si el dado es honesto en un lanzamiento el valor observado coincidirá con el número escogido al azar 1 de 6 veces, o sea con probabilidad 1/6, mientras que no coincidirán con probabilidad 5/6. Ahora bien, el dado se lanza tres veces y tenemos tres repeticiones del fenómeno binario "acertar" (éxito) o "no acertar" (fracaso) el número obtenido. Tenemos pues una variable binomial con n = 3 y p = 1/6.

$$P(X = x) = C(3,x)\cdot p^{x}\cdot(1-p)^{3-x}$$

para x = 0, 1, 2, 3.

b. La ganancia será positiva siempre que X = 1, 2, o 3. Como son sucesos excluyentes se deberá calcular:

$$P(X > 0) = P\big((X = 1)\cup(X = 2)\cup(X = 3)\big)$$

c. En este caso, al ser binomial, $E(X) = n\cdot p$. Sin embargo, si medimos las ganancias del juego, observaremos que la esperanza matemática es negativa:

$$E(G) = \sum_i g_i \cdot P(G = g_i) =$$
$$(-100) \cdot P(G = -100) + 100 \cdot P(G = 100) + 200 \cdot P(G = 200) + 300 \cdot P(G = 300) =$$
$$(-100) \cdot P(X = 0) + 100 \cdot P(X = 1) + 200 \cdot P(X = 2) + 300 \cdot P(X = 3) = -7{,}87$$

d. La varianza del juego, al ser binomial, se puede calcular por:

$$\mathrm{var}(X) = n \cdot p \cdot (1 - p)$$

Si queremos la varianza de las ganancias del juego, entonces la podemos calcular usando la expresión:

$$\mathrm{var}(G) = E(G^2) - E(G)^2$$

e. El precio "justo" desde la perspectiva del organizador del juego sería el valor q tal que $E(G) = q$, bajo el nuevo supuesto que $G = 0$ si $X = 0$. Una vez derivado el valor q, desde la perspectiva del apostante, el valor del juego se obtendría a partir de su utilidad esperada, descontando el "precio" q que ha de pagar para participar en el juego. En otras palabras, deberíamos obtener:

$$E(u(G - q))$$

Es fácil de ver que $q = 50$. Los premios posibles de este juego serían -50, 50, 150, 250 según $X = 0, 1, 2, 3$. Como disponemos de las probabilidades binomiales y una especificación de la utilidad, hallaríamos:

$$E(u(G - q)) =$$
$$u(-50) \cdot P(X = 0) + u(50) \cdot P(X = 1) + u(150) \cdot P(X = 2) + u(250) \cdot P(X = 3) =$$
$$= -0{,}714$$

A este precio "justo" para el organizador, el apostante no desearía participar pues obtendría un valor negativo.

61. (*) ***Casino Royale.*** Bond, James Bond, participa en un juego de azar consistente en lanzar una moneda "honesta" 8 veces. Si el resultado del lanzamiento es que aparecen 6, 7 u 8 caras entonces Bond, James Bond, obtiene un premio de 1.000 euros. En caso contrario Bond, James Bond, no obtiene nada.

   a. Calcular la probabilidad que Bond, James Bond, obtenga el premio.
   b. Calcular el premio esperado por Bond, James Bond.
   c. Evaluar la probabilidad que Bond, James Bond, gane tres veces el premio si el juego de azar descrito en el enunciado se repitiera 5 veces.

a. El suceso *A*: "Bond, James Bond, gana el premio de 1.000 euros" ocurrirá si salen 6, 7 u 8 caras. Vamos a definir el suceso "salir 6 caras" por *C*6, "salir 7 caras" por *C*7, etc. El suceso *A* no es otra cosa que la unión de *C*6, *C*7 y *C*8. Como son mutuamente excluyentes podemos aplicar la regla aditiva de la probabilidad para concluir:

$$P(A) = P(C6 \cup C7 \cup C8) = P(C6) + P(C7) + P(C8)$$

Ahora toca calcular estas probabilidades. Puesto que el lanzamiento de una moneda "honesta" 8 veces es un fenómeno aleatorio binomial con parámetros n = 8 y p = 0,5 tendremos:

$$P(C6) = C(8,6) \cdot (0,5)^6 \cdot (0,5)^{(8-6)} = 0,1094$$
$$P(C7) = C(8,7) \cdot (0,5)^7 \cdot (0,5)^{(8-7)} = 0,0313$$
$$P(C8) = C(8,8) \cdot (0,5)^8 \cdot (0,5)^{(8-8)} = 0,0039$$

Sumando hallamos que $P(A)$ = 0,1446.

b. Bond, James Bond, gana 1.000 euros con probabilidad 0,1446 y 0 euros con la probabilidad complementaria de 0,8554. Podemos pensar en que tenemos una variable aleatoria *X* que solo toma esos dos valores reales con sus respectivas probabilidades de forma que:

$$E(X) = 1.000 \cdot 0,1446 + 0 \cdot 0,8554 = 144,6$$

c. La probabilidad de ganar 3 veces si se repiten los lanzamientos en 5 ocasiones de nuevo es un fenómeno aleatorio binomial. La probabilidad de “éxito” en el primer lanzamiento de las monedas hemos visto que es $P(A) = 0{,}1446$. Para evaluar la probabilidad de tener 3 “éxitos” en 5 repeticiones podemos usar una variable binomial Z con parámetros $n = 5$ y $p = P(A)$. De esta manera:

$$P(Z=3) = C(5,3)\cdot P(A)^3(1-P(A))^{(5-3)} = 0{,}0221$$

62. Una empresa manufactura chips para ordenadores. De la experiencia acumulada se sabe que el 5% de los chips salen defectuosos de la línea de fabricación. Para el mercado, la empresa prepara cajas que contiene 10 chips. Existe un sistema de control de calidad sobre las cajas de forma que si en una de ellas hay un mínimo de 2 chips defectuosos, la caja se retira de la circulación.

a. Evaluar la probabilidad de tener que retirar una caja de chips.
b. Determinar el número esperado de chips defectuosos por caja y la varianza correspondiente.

a. Definiremos por X la variable aleatoria discreta que mide el número de chips defectuosos en cajas de 10 chips. Cada chip puede ser defectuoso (con probabilidad del 5%) o no serlo (con probabilidad del 95%). En consecuencia cada chip puede verse como una variable binaria y cada caja puede verse como 10 repeticiones de la variable binaria. Esto corresponde a una variable binomial con $n = 10$ y $p = 0{,}05$. Una caja se retira si $X \geq 2$. La probabilidad que queremos evaluar es $P(X\geq 2)$. La forma más rápida es usar la regla del complementario y aplicar la ley binomial:

$$P(X\geq 2) = 1- P(X<2) = 1- P(X=0) - P(X=1)$$

Usando la binomial obtenemos:

$$P(X=0) = C(10,0)\cdot 0{,}05^0\cdot 0{,}95^{10} = 0{,}5987$$
$$P(X=1) = C(10,1)\cdot 0{,}05^1\cdot 0{,}95^{9} = 0{,}3151$$

De aquí se sigue:

$$P(X\geq 2) = 1- 0{,}5897- 0{,}3151= 0{,}0862$$

b. Ahora usamos directamente las expresiones para la esperanza y la varianza de una variable X que es binomial con parámetros n y p:

$$E(X) = n \cdot p = 10 \cdot 0{,}05 = 0{,}50$$
$$\mathrm{Var}(X) = n \cdot p \cdot (1-p) = 10 \cdot 0{,}05 \cdot 0{,}95 = 0{,}48$$

63. (*) La asignatura de Estadística acostumbra a ser un "hueso duro de roer". La experiencia acumulada muestra que para un alumno típico del curso la probabilidad de superar satisfactoriamente el primer examen final es del 15%. Considerar un grupo de 5 alumnos que se enfrentan a su primer examen final y evaluar las siguientes probabilidades:

a. Ninguno de los alumnos supera el examen.
b. Uno de ellos lo supera.
c. Lo supera al menos uno de ellos.
d. Si el grupo bajo consideración fuese de 100 alumnos, evaluar la probabilidad que superen el primer final 20 de ellos.

a. Se supera la materia (un "éxito") con $p = 0{,}15$. Se trata de ver el número de éxitos en n repeticiones (con $n = 5$, o $n = 100$) y tenemos en consecuencia una variable binomial. En el primer caso (con $n = 5$) queremos evaluar $P(X=0)$:

$$P(X=0) = C(n,0) \cdot p^0 \cdot (1-p)^n = C(5,0) \cdot (0{,}15)^0 \cdot (0{,}85)^5 = 44{,}37\%$$

b. Para $n = 5$ buscamos $P(X = 1)$

$$P(X=1) = C(n,1) \cdot p^1 \cdot (1-p)^{n-1} = C(5,1) \cdot (0{,}15)^1 \cdot (0{,}85)^4 = 39{,}15\%$$

c. Para $n = 5$ buscamos $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$ con esta última probabilidad sacada del apartado a. Así, $P(X \geq 1) = 55{,}63\%$

d. Para $n = 100$ buscamos $P(X = 20)$. Usando la binomial sería:

$$P(X=20) = C(n,20) \cdot p^{20} \cdot (1-p)^{n-20} = C(100,20) \cdot (0{,}15)^{20} \cdot (0{,}85)^{80}$$

Podemos ir un poco más rápidos aproximando la binomial por una variable de Poisson con $\lambda = n \cdot p = 100 \cdot 0{,}15 = 15$ (obsérvese que la aproximación es posible porque n es grande y p es pequeña). Así pues:

$$P(X=20) = \frac{e^{-\lambda} \cdot \lambda^{20}}{20!} = 4{,}18\%$$

A título comparativo, la probabilidad binomial evaluada usando la fórmula binomial genera una probabilidad de 4,02%. A efectos prácticos la aproximación de Poisson ofrece una estimación suficientemente buena.

64. Se han realizado una serie de verificaciones sobre la calidad de unos componentes usados en la fabricación de coches y se ha detectado que el 4% de esos componentes presenta algún tipo de problema que los hace inservibles. Se han instalado 150 de estos componentes.

a. Evaluar la probabilidad que 4 de esos componentes instalados sean defectuosos.
b. Evaluar la probabilidad que hayan entre 4 y 6 defectuosos.
c. Comentar sobre el valor de la esperanza y la varianza.

a. La estructura del problema responde a una variable binomial con n = 150 y p = 0,04. Como la probabilidad *p* es relativamente reducida y el número de repeticiones *n* es relativamente grande es más sencillo y rápido usar la aproximación de Poisson a una variable binomial. El parámetro de Poisson lo calculamos como:

$$\lambda = n \cdot p = 150 \cdot 0{,}04 = 6$$

De aquí usamos la ley de Poisson para determinar la probabilidad de 4 componentes en mal estado:

$$P(X = 4) = \frac{e^{-\lambda} \cdot \lambda^4}{4!} = 0{,}1339$$

b. En este caso necesitamos evaluar la probabilidad en un intervalo discreto:

$$P(4 \leq X \leq 6) = \sum_{x=4}^{6} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = 0{,}4451$$

c. Para una variable de Poisson, la esperanza y la varianza coinciden con el parámetro λ.

65. (*) Una estación de la Inspección Técnica de Vehículos (ITV) ha revisado 48.000 vehículos durante todo el año y ha detectado que 36 de estos vehículos tenían los cinturones de seguridad defectuosos. El número de coches revisados cada mes es aproximadamente el mismo.

   a. Calcular la probabilidad que en un mes dado la estación haya detectado exactamente 2 coches con los cinturones defectuosos.
   b. Calcular la probabilidad que en un mes dado la estación haya detectado un mínimo de 8 vehículos con cinturones defectuosos.

   a. Cada uno de los vehículos puede tener los cinturones de seguridad defectuosos o no. La probabilidad de observar un cinturón defectuoso la podemos evaluar en términos de los casos observado (36) sobre el total de examinados (48.000). Por consiguiente p = 36/48.000 = 0,00075. El número de vehículos revisados por mes es de 48.000/12 = 4,000. Tenemos pues n = 4.000 repeticiones del experimento binario cinturón defectuoso/no defectuoso que, además, pueden considerarse como independientes. La situación se corresponde con la de una variable binomial con parámetros n = 4.000 y p = 0,00075. La probabilidad buscada se puede evaluar a partir de:

$$P(X=2) = C(n,2)\cdot p^2 \cdot (1-p)^{n-2} =$$
$$= \frac{4.000!}{2!\cdot(4.000-2)!}\cdot(0{,}00075)^2\cdot(0{,}99925)^{(4.000-2)} = 0{,}2241$$

   b. Para la misma variable binomial ahora hemos de evaluar $P(X \geq 8)$. Un primer paso para simplificar el cálculo es pasar al suceso complementario $(X < 8) = (X \leq 7)$ y evaluar su probabilidad. Un segundo paso simplificador resulta de percibir que esta situación se puede aproximar por la variable de Poisson pues el número de repeticiones es elevado (n = 4.000) y la probabilidad del suceso binario es pequeña (p = 0,00075). Construimos pues el parámetro de la variable de Poisson que aproxima a la variable binomial: $\lambda = n\cdot p = (4.000)\cdot(0{,}00075) = 3$ y usando la tabla de probabilidades acumuladas de Poisson, para el valor λ = 3, hallamos que $P(X \leq 7) = 0{,}988$. Finalmente $P(X \geq 8) = 1 - P(X \leq 7) = 0{,}012$. La probabilidad de detectar al menos 8 vehículos por mes con cinturón defectuoso es francamente baja, solo del 1,2%.

66. (*) Un examen contiene una parte con preguntas de teoría en formato test que deben señalarse como verdaderas o falsas. Para superar la prueba el profesor ha decidido que se deben contestar correctamente al menos 7 de las 10 preguntas.

   a. Si un estudiante no tiene ni idea de la materia y contesta al azar (lanzando por ejemplo una moneda honesta al aire), evaluar la probabilidad que supere el examen.
   b. Estudiar la materia tiene la virtud que hace aumentar la probabilidad de detectar correctamente la respuesta. Supongamos, en particular, que estudiar aumenta la probabilidad de acierto en un 40%. Volver a evaluar la probabilidad que un estudiante que haya estudiado supere el examen.

   a. Superar el examen supone acertar 7 o más de las respuestas. En el caso concreto de 7 hay $C(10,7)$ maneras distintas de acertar al azar 7 de las 10 preguntas. Cada una de estas maneras tiene una probabilidad evaluable en $p^7 \cdot (1-p)^3$ donde p es la probabilidad de acertar al azar (p = 0,5). Tenemos pues una variable binomial. Si aplicamos el mismo argumento a acertar 8, 9 o 10 respuestas tendremos la probabilidad buscada:

$$P = C(10,7)\cdot p^7 \cdot (1-p)^3 + C(10,8)\cdot p^8 \cdot (1-p)^2 + \\ C(10,9)\cdot p^9 \cdot (1-p)^1 + C(10,10)\cdot p^{10} \cdot (1-p)^0 = 0{,}172$$

   b. Si con el estudio la probabilidad aumenta en un 40%, ahora tendríamos p = 0,7. Sustituyendo en la expresión anterior hallaríamos P = 0,649 y estudiar casi cuadriplica la probabilidad de superar la prueba. ¡No está nada mal!

67. El gerente de un restaurante con 20 mesas que solo ofrece cenas mediante reservas sabe, por experiencia, que el 15% de los grupos que reservan mesa después no se presentan a la hora de la cena. Para evitar que queden mesas sin ocupar, un día en concreto el restaurante acepta 23 reservas:

   a. ¿Cuál es el número esperado de grupos con reserva que se presentarán este día en el restaurante para cenar?
   b. ¿Cuál es la probabilidad de que todos los grupos con reserva que se presenten este día en el restaurante puedan ser servidos, es decir, puedan sentarse en una mesa?
   c. Poco antes de empezar la cena el gerente observa que las primeras 19 reservas se han presentado, no ha fallado ninguna. ¿Cuál es la probabilidad de que todos los grupos con reserva que lleguen después puedan sentarse en una mesa?

   a. Vamos por pasos. La variable aleatoria

$$X = \text{“Número de grupos con reserva que se presentan a la cena”}$$

sigue una distribución binomial de parámetros $n = 23$ y $p = 0{,}85$. Esto es así porque se producen 23 repeticiones (23 reservas) de un suceso que puede resultar en "presentarse" o "no presentarse" con probabilidades 0,85 y 0,15 respectivamente. Por tanto tenemos:

$$X \sim B(23;0{,}85)$$

Por tanto, si nos preguntan por el valor esperado de X podemos usar la expresión de la esperanza para una variable binomial

$$E(X) = n{\cdot}p = 23{\cdot}0{,}85 = 19{,}55$$

b. Aquí nos preguntan por la probabilidad que se presenten 20 grupos como máximo. Esta es la única manera que todos puedan tener mesa. Es decir:

$$P(X \leq 20) = 1 - P(X > 20) = 1 - [P(X=21) + P(X=22) + P(X=23))]$$

Si calculamos cada una de estas probabilidades usando la fórmula binomial tendremos:

$$P(X = 21) = C(23,21)\cdot(0{,}85)^{21}\cdot(0{,}15)^{2} = 0{,}187$$
$$P(X = 22) = C(23,22)\cdot(0{,}85)^{22}\cdot(0{,}15)^{1} = 0{,}097$$
$$P(X = 23) = C(23,23)\cdot(0{,}85)^{23}(0{,}15)^{0} = 0{,}024$$

Por tanto,

$$P(X \leq 20) = 1 - P(X>20) = 1 - [P(X=21) + P(X=22) + P(X=23))] =$$
$$= 1 - [0{,}187+0{,}097+0{,}024] = 1 - 0{,}31 = 0{,}69$$

c. Definamos una nueva variable aleatoria Y como:

Y = "Número de grupos con reserva que se presentan a la cena cuando los primeros 19 grupos con reserva se han presentado"

Está claro que Y solo puede tomar los valores {0,1,2,3,4}, ya que no pueden presentarse más de 4 grupos. Por tanto, tenemos que Y es una variable binomial con n = 4 y p = 0,85:

$$Y \sim B(4;0{,}85)$$

Nos piden la probabilidad que, como máximo, se presente 1 grupo más. Esta es la única manera que todos puedan tener mesa:

$$P(Y \leq 1) = P(Y=0) + P(Y=1)$$

Si calculamos cada una de estas probabilidades tenemos:

$$P(Y=0) = \binom{4}{0}(0{,}85)^0(0{,}15)^4 = 0{,}0005$$

$$P(Y=1) = \binom{4}{1}(0{,}85)^1(0{,}15)^3 = 0{,}0115$$

Por tanto,

$$P(Y \leq 1) = P(Y=0) + P(Y=1) = 0{,}0005 + 0{,}0115 = 0{,}012$$

68. Suponga que usted es un economista empleado por la directiva de un famoso club de fútbol, el Barça, para averiguar ciertas probabilidades sobre las características de la afición. En particular, consideraremos una muestra de 10 individuos escogidos al azar de la población de Cataluña y queremos averiguar las probabilidades de:

    1. 6 de los individuos son aficionados del Barça
    2. Por lo menos 6 son del Barça
    3. Más de 6 son del Barça
    4. Ninguno es del Barça
    5. Todos son del Barça

    a. Discuta qué variable aleatoria (o qué distribución de probabilidad) podría utilizar para ayudarle a evaluar esas probabilidades.
    b. Póngase manos a la obra y haga los cálculos. Piense que este problema requiere cierta investigación por su parte más allá del trabajo de aula pues, de una forma u otra, tiene que obtener un valor de un parámetro fundamental para obtener su respuesta (¿cuál?). Explique qué método(s) práctico(s) podría usar para obtenerlo.

a. Cada habitante puede ser aficionado del Barça o no serlo (porque no le interesa el fútbol, es de otro equipo, etc.). Podemos considerar que la situación responde a un fenómeno binario (*to be or not to be*, *that is the question*) con probabilidad p de ser aficionado del Barça y $(1 - p)$ de no serlo.

   Al escoger 10 individuos al azar, podemos considerar que tenemos 10 repeticiones del fenómeno binario *to be or not to be*. En este caso, el número de individuos entre los 10 que son aficionados se puede medir por una variable binomial X con parámetros $n = 10$ y p. Que exactamente 6 de los 10 sean aficionados del Barça tiene probabilidad que se puede evaluar por:

$$P(X = 6) = C(10,6) \cdot p^6 \cdot (1 - p)^{10-6}$$

   Que por lo menos 6 sean de Barça tendría probabilidad:

$$P(X \geq 6) = \sum_{k=6}^{10} C(10,k) \cdot p^k \cdot (1 - p)^{10-k}$$

   Si nos interesa la probabilidad que hayan más de 6, tendríamos que evaluar:

$$P(X > 6) = \sum_{k=7}^{10} C(10,k) \cdot p^k \cdot (1 - p)^{10-k}$$

   Que nadie sea del Barça tendría probabilidad evaluable por:

$$P(X = 0) = C(10,0) \cdot p^0 \cdot (1 - p)^{10-0}$$

   Finalmente, la probabilidad que los 10 sean aficionados se podría evaluar por:

$$P(X = 10) = C(10,10) \cdot p^{10} \cdot (1 - p)^{10-10}$$

b. Necesitaríamos saber el valor de p, esto es, el porcentaje de la población de catalanes que se sienten aficionados del Barça. Para obtener esta cifra podríamos explorar si el club ha hecho una encuesta y nos facilita los datos. O podríamos hacer una pequeña encuesta nosotros preguntando, por ejemplo, a 20 individuos que pasean por la calle. Se trataría de que la encuesta fuese realmente realizada al azar, e incluso si no es una encuesta sobre una muestra representativa, ya nos daría una primera aproximación. A título de ejemplo, si supiéramos que $p = 0,6$, las probabilidades previas serían, respectivamente, las siguientes:

$$P(X=6)=0{,}2508$$
$$P(X\geq 6)=0{,}6331$$
$$P(X>6)=0{,}3823$$
$$P(X=0)=0{,}0001$$
$$P(X=10)=0{,}0060$$

69. Durante una campaña promocional, un concesionario de coches garantiza a los compradores de coches que si no están satisfechos con el modelo comprado lo pueden devolver en el plazo de una semana y se les reintegrará todo el precio. Por cada coche vendido, el concesionario obtiene un beneficio de 2.000 euros y en caso de devolución se incurre en un coste de 250 euros. La empresa consigue vender 100 coches durante esta campaña, pero estima que un 15% de los compradores devolverán el coche.

   a. Discuta qué variable aleatoria X describe el número de coches devueltos x.
   b. Calcule la media y la desviación estándar del número de coches devueltos.
   c. Describa la distribución de probabilidad de la variable aleatoria B que mide el beneficio del concesionario en función de los coches devueltos.
   d. Calcule el beneficio medio y su desviación estándar sin usar la distribución de probabilidad de B.

a. Dado que cada coche puede ser devuelto con una probabilidad p = 0,15 y que la empresa ha vendido 100 coches, tenemos que X = “número de coches vendidos y devueltos” sigue una distribución binomial de parámetros n = 100 y p = 0,15

$$X \sim B(100;0{,}15)$$

b. Por las características de la distribución binomial sabemos que

$$E(X)=n\cdot p=100\cdot 0{,}15=15$$

Es decir, se espera que 15 coches sean devueltos. De la misma manera, usando la expresión para la varianza de una variable binomial tendríamos:

$$\operatorname{var}(X)=n\cdot p\cdot(1-p)=100\cdot 0{,}15\cdot 0{,}85=12{,}75$$
$$\sigma_X=\sqrt{\operatorname{var}(X)}=\sqrt{12{,}75}=3{,}57$$

c. El beneficio será:

$$B = \text{Ingresos} - \text{Costes}$$
$$B = 100 \cdot 2.000 - X \cdot (250 + 2.000) = 200.000 - 2.250 \cdot X$$

Dado que $X \in \{0,1,2,\ldots,100\}$ tendremos que:

$$B \in \{-25.000, -22.750, \ldots, 197.750, 200.000\}$$

Finalmente, a partir de la expresión del beneficio de la empresa en función de X tenemos:

$$P(B = b) = P(X = \frac{200.000 - b}{2.250})$$

d. Utilizando las propiedades de la esperanza,

$$E(B) = E(200.000 - 2.250 \cdot X) = 200.000 - 2.250 \cdot E(X)$$
$$E(B) = 200.000 - 2.250 \cdot 15 = 166.250$$

De manera parecida, por las propiedades de la varianza,

$$\text{var}(B) = \text{var}(200.000 - 2.250 \cdot X) = 0 + 2.250^2 \cdot \text{var}(X)$$
$$\text{var}(B) = 5.062.500 \cdot 12,75 = 6.4546.875$$

Por tanto, la desviación estándar será:

$$\sigma_B = \sqrt{\text{var}(B)} = \sqrt{64546875} = 8.034,11$$

70. ***Overbooking en el Prat***. Las líneas aéreas acostumbran a vender (¡y la normativa lo permite!) más billetes para un viaje que la capacidad disponible de los aviones permite. Suponga que el vuelo Barcelona-París tiene capacidad para 110 viajeros, pero la compañía vende 120 billetes. En la práctica se sabe que no todos los viajeros con billete aparecen el día del vuelo y que el porcentaje de viajeros que no se presenta es del 5%, aproximadamente.

a. Discutir qué tipo de variable aleatoria describe este tipo de fenómeno.
b. Calcular el número esperado de viajeros del mencionado vuelo. Calcular también la desviación estándar.
c. Calcular la probabilidad que todos los viajeros que se presentan el día del vuelo puedan hacer el viaje que planeaban. ¿Qué dificultad práctica encuentra para hallar su respuesta? Indique un camino alternativo para resolver el problema.
d. Suponer que la compañía desea que la probabilidad de que todos los viajeros que se presentan puedan hacer el vuelo es del 50%. Averiguar cuántos billetes por encima de la capacidad del avión debería vender la línea aérea para cumplir con este requisito.

a. Cada uno de los viajeros puede acabar no apareciendo el día previsto del viaje por razones varias. Usaremos X para denotar la variable aleatoria que mide el número de viajeros que aparecen el día del viaje. Puesto que se venden 120 billetes, tendremos que hay n = 120 repeticiones de la situación binaria "aparecer" o "no aparecer", sabiendo además que la probabilidad de aparecer se estima que es en la práctica del 95%. Tenemos pues una variable binomial con n = 120 y p = 0,95.

b. Al ser X una variable binomial la respuesta es inmediata. La esperanza matemática es $E(X) = n \cdot p = 120 \cdot 0{,}95 = 114$, mientras que la desviación estándar la obtenemos de la varianza binomial:

$$\sigma_X = \sqrt{\text{var}(X)} = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{120 \cdot 0{,}05 \cdot 0{,}95} = 2{,}39$$

c. Puesto que X mide el número de viajeros que aparecen el día del vuelo, el suceso se puede representar por ($X \le 110$) y la probabilidad en cuestión se puede evaluar usando la variable binomial a partir de $P(X \le 110)$. Si hacemos que Y mida el número de pasajeros que no aparecen el día del vuelo tendremos que se habrá de cumplir que $X+Y$ = 120, de manera que:

$$P(X \le 110) = P(120 - Y \le 110) = P(Y \ge 10)$$

siendo Y también una variable binomial para la que tenemos que n = 120 y q = 0,05. Si aproximamos la variable binomial con una variable de Poisson (podemos hacerlo satisfactoriamente pues n es relativamente grande y q pequeña) con parámetro $\lambda = n \cdot q = 120 \cdot 0{,}05 = 6$, podemos evaluar la probabilidad a partir de:

$$P(Y \ge 10) = 1 - P(Y < 10) = 1 - P(Y \le 9) = 1 - 0.916 = 0{,}084 = 8{,}4\%$$

Tal cifra se puede obtener asimismo de las tablas de probabilidad acumulada de Poisson.

d. Debemos hallar el número de billetes vendidos b tal que $X + Y = b$ y además se verifique que $P(X \leq 110) = 0{,}50$ donde X es ahora binomial con parámetros b y p. Siguiendo el mismo razonamiento anterior:

$$P(X \leq 110) = P(b - Y \leq 110) = P(Y \geq b - 110) = P(Y \geq e)$$

donde e = b-110 es el número de billetes en exceso. Puesto que,

$$P(Y \geq e) = 1 - P(Y < e) = P(Y \leq e - 1) = 0{,}50$$

y la variable Y se puede de nuevo aproximar por una variable de Poisson con $\lambda = 6$, de las tablas de la variable de Poisson podemos ver que e–1 estará entre 5 y 6, o sea que e estará entre 6 y 7 y en consecuencia la compaña aérea debería vender un número de billetes b entre 116 y 117, si desea que la probabilidad que los viajeros que se presentan puedan realizar el vuelo sea del 50%.

71. (*) Según una información publicada en el diario *La Vanguardia*, cada 5 segundos una persona se contagia del virus de la sida (o sea, 12 cada minuto; fíjese que equivale a ¡¡más de seis millones de personas al año!!). Con estos datos, calcular la probabilidad que en un día se contagien más de 35.000 personas.

Una primera observación es que en un intervalo temporal de 5 segundos se produce un contagio. La ocurrencia de un contagio en ese intervalo es un suceso infrecuente. En consecuencia puede interpretarse como un suceso regido por la variable de Poisson. Podemos pues atribuir el parámetro λ = 1, que describe los contagios por intervalo en promedio, a la variable de Poisson. Para poder contestar la pregunta hemos de modificar el intervalo temporal. Por ejemplo, un contagio cada 5 segundos comporta un total de 5 × 12 = contagios en un minuto (pues hay 12 intervalos de 5 segundos en un minuto), o equivalentemente a 720 contagios por hora (1 hora = 60 minutos) o 17.280 contagios por día (1 día = 24 horas). Más de 35.000 contagios por día, haciendo el camino inverso, nos dice que en un intervalo de 5 segundos se deberían producir más de 2,02 contagios (35.000/24/60/12 = 2,02).

La probabilidad buscada, en el intervalo básico de 5 segundos, será evaluada por Poisson con λ = 1 usando

$$P(X > 2{,}02) \approx P(X > 2) = 1 - P(X \leq 2) =$$
$$1 - P(X = 0) - P(X = 1) - P(X = 2) = 0{,}323$$

En definitiva, la probabilidad de que ocurran más de 35.000 contagios por día es aproximadamente del 32,3%.

72. ***La ruta del bacalao***. Una empresa pesquera dedicada a la pesca del bacalao utiliza un avión de exploración y búsqueda de bancos de dicho animal marino. En el Atlántico Norte se sabe que en promedio existe un banco de bacalao por cada 100.000 km cuadrados de mar. En un día normal de trabajo el avión puede recorrer 1.000 km de distancia y los sensores de detección pueden abarcar hasta 5 km a cada lado de la trayectoria del avión.

    a. Averiguar la probabilidad de detectar un banco de bacalao en tres días de exploración.
    b. Repetir para un período de seis días.
    c. Averiguar cuántos días de búsqueda serían necesarios para que la probabilidad de hallar por lo menos un banco de peces sea del 95%.

    a. La probabilidad de detectar un banco de peces en 1 km^2 es de p = 1/100.000. En tres días de trabajo se pueden explorar 30.000 km^2 de superficie marina. Esta cifra se obtiene a partir del dato que en un día el avión recorre 1.000 km lineales y a cada lado de la trayectoria se pueden escanear 5 km. En tres días tendremos, en consecuencia, un total de 1.000 km × (5 km × 2) × 3 = 30.000 km^2 en los que detectar la presencia de bancos de peces.

    Cosas que sabemos: i) en cada km^2 hay una probabilidad p de detectar un banco, ii) hay 30.000 km^2 para analizar con el escáner y en cada uno de ellos puede haber (o no) un banco, iii) una variable binomial es apropiada para medir el número de detecciones positivas (éxitos) y en este caso n = 30.000 y p = 1/100.000, iv) puesto que n es grande y p es pequeña, podemos aproximar la variable binomial con una variable aleatoria de Poisson con parámetro $\lambda = n \cdot p = 0{,}3$.

    Para responder a la pregunta formulada sobre la probabilidad de detección podemos considerar dos casos: el primero es detectar exactamente un banco, es decir, $P(X = 1)$, o bien detectar por lo menos un banco y en este caso necesitaríamos evaluar $P(X \geq 1)$. Usando la fórmula de Poisson tendríamos:

$$P(X = 1) = \frac{e^{-\lambda} \cdot \lambda^1}{1!} = 0{,}222$$

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{e^{-\lambda} \cdot \lambda^0}{0!} = 0{,}259$$

    b. Este apartado es similar al previo, con la sustitución del número de días de exploración. Dejamos los detalles de cálculo para el lector.

c. En este caso necesitamos determinar el número de días de exploración que garantice que la probabilidad de detección es del 95%. Partiremos de usar la probabilidad de detectar al menos un banco de peces de forma que de

$$P(X \geq 1) = 1 - P(X = 0) = 0{,}95$$

deducimos que

$$P(X = 0) = 0{,}05$$

Usando la fórmula de Poisson para este valor de probabilidad:

$$P(X = 0) = \frac{e^{-\lambda} \cdot \lambda^0}{0!} = e^{-\lambda} = 0{,}05$$

se sigue que el parámetro de la expresión es $\lambda \approx 3$. Puesto que $\lambda = n \cdot p$ y n es el número de días multiplicado por los km^2 por día, es decir, $n = d \times 10.000$ tendremos que:

$$\lambda = n \cdot p = d \cdot (10.000) \cdot (1/100.000) \approx 3$$

de forma que observamos que el número de días necesarios es $d \approx 30$.

73. (*) Una empresa de servicios tiene un servicio de atención telefónica a los clientes. La experiencia prueba que en promedio se atiende a cuatro clientes cada diez minutos. El servicio puede atender hasta seis clientes cada diez minutos sin incurrir en colas de espera (que siempre son molestas).

a. Comentar qué tipo de variable aleatoria puede describir este tipo de fenómenos.
b. Averiguar la probabilidad que el servicio de atención al cliente se sature. Suponer a continuación que la empresa considera que el servicio es exitoso si ningún cliente es sujeto a ninguna espera.
c. Averiguar la probabilidad que en una hora se produzcan dos periodos en los que hay esperas.
d. Estimar el número posible de clientes satisfechos durante una hora. ¿Es posible saber el número de clientes insatisfechos durante una hora?

a. La llegada de peticiones de asistencia es un intervalo de tiempo es una situación que puede analizarse con una variable de Poisson, pues el número de ocurrencias es escaso dentro del intervalo.

b. Si hacemos que X mida el número de peticiones de asistencia, la saturación ocurrirá cuando estas excedan a 6, i. e. necesitamos evaluar la probabilidad

$$P(X > 6) = 1 - P(X \leq 6)$$

Como en promedio se reciben 4 peticiones de asistencia por intervalo, tenemos que el parámetro de la variable de Poisson será $\lambda = 4$. Bien por cálculo directo, bien usando las tablas de probabilidad acumulada de Poisson, obtenemos que $P(X \leq 6) = 0{,}889$ y de aquí que la probabilidad buscada será de $P(X > 6) = 0{,}111$.

c. Una hora puede verse como una cadena de 6 intervalos de 10 minutos. En cada uno de estos intervalos puede darse saturación ("fracaso") con probabilidad 0,111 o no-saturación ("éxito") con probabilidad 0,889. Esta situación puede analizarse usando una variable aleatoria Z, binomial con $n = 6$ y $p = 0{,}111$, de manera que el suceso "dos periodos con saturación" tendrá probabilidad

$$P(Z = 2) = C(6,2) \cdot (0{,}111)^2 \cdot (0{,}889)^4 = 0{,}1154 = 11{,}54\%$$

d. Los clientes estarán satisfechos si no se les hace esperar cuando llaman para obtener asistencia. En una hora el número de periodos con saturación se puede calcular como el valor esperado de Z:

$$E(Z) = n \cdot p = 6 \cdot 0{,}111 = 0{,}666$$

y en consecuencia, al haber 6 periodos de 10 minutos en una hora, el número de exitosos (sin saturación) será de 6 – 0,666 = 5,334. Como en cada periodo hay en promedio 4 clientes satisfechos (recordemos que $\lambda = 4$), tendremos alrededor de 21 clientes satisfechos por hora ($4 \cdot 5{,}334 = 21{,}336 \approx 21$).

Variables aleatorias continuas

74. Suponga que un vuelo de Madrid a Barcelona tiene prevista su llegada a las 10:30 de la mañana. Sin embargo, el avión igual puede llegar antes que después de la hora prevista. Asumiendo que la llegada del avión se distribuye uniformemente y que el avión puede llegar en cualquier momento entre las 10:15 y las 10:50 horas,

   a. Calcular la probabilidad que el avión no llegue con retraso.
   b. Calcular la probabilidad que un viajero del vuelo de Madrid que ha de hacer transbordo a un vuelo hacia Roma que sale a las 10:40 horas pierda la conexión.

   a. En primer lugar resulta conveniente convertir los datos que se ofrecen en formato horario (hh:mm) al formato decimal de los números reales (recordar que una variable aleatoria es un variable real). El factor de conversión de minutos a decimal es 1/60. Así tenemos

$$10:15 \rightarrow 15 \text{ minutos} \cdot \frac{1}{60} = 0{,}25 \rightarrow 10{,}25$$
$$10:30 \rightarrow 30 \text{ minutos} \cdot \frac{1}{60} = 0{,}50 \rightarrow 10{,}50$$
$$10:40 \rightarrow 40 \text{ minutos} \cdot \frac{1}{60} = 0{,}67 \rightarrow 10{,}67$$
$$10:50 \rightarrow 50 \text{ minutos} \cdot \frac{1}{60} = 0{,}83 \rightarrow 10{,}83$$

En segundo lugar observamos que la variable X que representa la hora de llegada del avión sigue una distribución uniforme entre los valores 10,25 y 10,83. Por tanto, su función de densidad f_X vendrá dada por:

$$f_X = \begin{matrix} \frac{1}{10{,}83 - 10{,}25} & 10{,}25 \le x \le 10{,}83 \\ 0 & \text{en caso contrario} \end{matrix} = \begin{matrix} \frac{1}{0{,}58} & 10{,}25 \le x \le 10{,}83 \\ 0 & \text{en caso contrario} \end{matrix}$$

Por tanto, si la variable X representa la hora de llegada del avión, la probabilidad de que el avión no llegue con retraso es igual a la probabilidad de que el avión llegue entre las 10:15 (10,25, en formato decimal) y las 10:30 (10,50 en formato decimal). O sea

$$P(10{,}15 \le X \le 10{,}50).$$

Calculamos,

$$P(10,25 \leq X \leq 10,50) = \int_{10,25}^{10,50} \frac{1}{0.58} dx = \left| \frac{1}{0.58} x \right|_{10,25}^{10,50} =$$

$$= \frac{10,50}{0,58} - \frac{10,25}{0,58} = 0,43$$

de forma que la probabilidad que el avión no llegue con retraso es del 43%.

b. El viajero perderá la conexión a Roma si el avión llega después de las 10:40 (10,67, en formato decimal). Por tanto, la probabilidad de que pierda la conexión es igual a la probabilidad que el avión llegue entre las 10:40 (10,67) y las 10:50 (10,83):

$$P(10,67 \leq X \leq 10,83)$$

Calculamos,

$$P(10,67 \leq X \leq 10,83) = \int_{10,67}^{10,83} \frac{1}{0.58} dx = \left| \frac{1}{0.58} x \right|_{10,67}^{10,83} =$$

$$= \frac{10,83}{0,58} - \frac{10,67}{0,58} = 0,27$$

La probabilidad de perder la conexión es casi del 30%, una cifra no despreciable para que el pasajero la tenga en cuenta en sus planes de viaje.

75. ***España va bien***. Suponga que un grupo de economistas le asegura que la tasa de inflación el próximo año no estará por encima del 5% ni por debajo del 2%. Si todos los valores intermedios son igualmente probables, averiguar:

a. La probabilidad que la inflación esté por encima del 4%.
b. La probabilidad que la inflación esté entre el 3% y el 4%.
c. La probabilidad que la inflación esté por debajo del 3%.
d. La inflación esperada y su dispersión en el próximo año.

a. Claramente tenemos que X = "Tasa de inflación en el próximo año" sigue las pautas de una variable uniforme entre los valores 2 y 5 (ambos expresados en porcentaje):

$$X \sim U[2,5]$$

Es decir, tenemos que:

$$f_X(x) = \begin{cases} \dfrac{1}{5-2} & x \in [2,5] \\ 0 & \text{en caso contrario} \end{cases} = \begin{cases} \dfrac{1}{3} & x \in [2,5] \\ 0 & \text{en caso contrario} \end{cases}$$

Por tanto, la probabilidad que la inflación esté por encima del 4% corresponde a $P(4 \le X \le 5)$. Calculando,

$$P(4 \le X \le 5) = \int_4^5 \frac{1}{3} dx = \left| \frac{1}{3} x \right|_4^5 = (\frac{1}{3}5 - \frac{1}{3}4) = \frac{1}{3}$$

Existe por tanto una probabilidad del 33,37% de que la inflación supere el 4% el próximo año.

b. La probabilidad que la inflación esté entre el 3% y el 4% corresponde a $P(3 \le X \le 4)$. Calculando,

$$P(3 \le X \le 4) = \int_3^4 \frac{1}{3} dx = \left| \frac{1}{3} x \right|_3^4 = (\frac{1}{3}4 - \frac{1}{3}3) = \frac{1}{3}$$

Existe de nuevo una probabilidad del 33,37% de que la inflación esté entre el 3% y el 4% el próximo año.

c. La probabilidad que la inflación esté por debajo del 3% corresponde a $P(2 \le X \le 3)$. Calculando,

$$P(2 \le X \le 3) = \int_2^3 \frac{1}{3} dx = \left| \frac{1}{3} x \right|_2^3 = (\frac{1}{3}3 - \frac{1}{3}2) = \frac{1}{3}$$

Una vez más, existe una probabilidad del 33,37% de que la inflación esté por debajo del 3% el próximo año.

d. Por tratarse de una distribución uniforme sabemos que,

$$E(X)=\frac{b+a}{2}=\frac{5+2}{2}=3{,}5$$
$$\operatorname{var}(X)=\frac{(b-a)^2}{12}=\frac{(5-2)^2}{12}=\frac{3}{4}$$
$$\sigma_X=\sqrt{\frac{3}{4}}=\frac{\sqrt{3}}{2}\approx 0.866$$

Así, se espera una inflación del 3,5% para el próximo año, con una desviación típica de 0,866%.

76. Utilizar la tabla de probabilidades para la distribución normal estandarizada para obtener las siguientes probabilidades (acompañar cada cálculo con el gráfico correspondiente):

a. $P(0\le Z\le 1)$
b. $P(1\le Z\le 2)$
c. $P(-0.5\le Z\le -0.2)$
d. $P(-1\le Z\le 2)$
e. $P(-1\le Z\le \infty)$
f. $P(-\infty\le Z\le 1)$
g. $P(|Z|\le 1)$

a. Por conveniencia, representamos con $\Phi(z)$ el valor que se encuentra en las tablas de la variable normal estandarizada y que corresponde a:

$$\Phi(z)=P(Z\le z)$$

donde $Z\sim N(0,1)$

Así,

$$P(0\le Z\le 1)=P(Z\le 1)-P(Z\le 0)=\Phi(1)-\Phi(0)$$

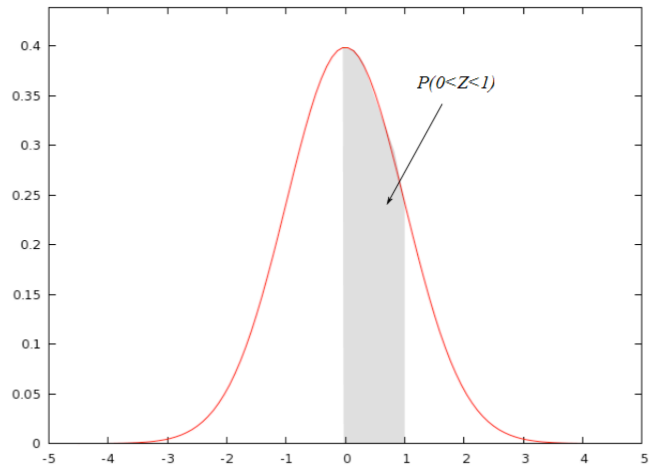
En las tablas de la normal estándar encontramos:

$$\Phi(1)=0{,}8413$$
$$\Phi(0)=0{,}5$$

Por tanto,

$$P(0 \leq Z \leq 1) = \Phi(1) - \Phi(0) = 0{,}8413 - 0{,}4 = 0{,}3413$$

Gráficamente:



b. Obtendremos ahora,

$$P(1 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq 1) = \Phi(2) - \Phi(1)$$

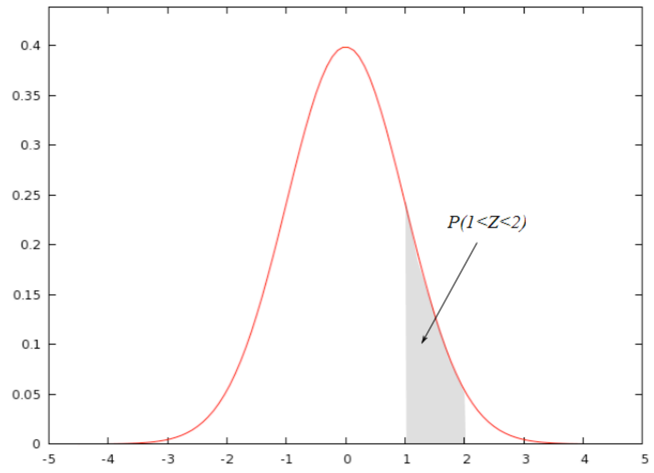
En las tablas de la normal estándar encontramos:

$$\Phi(1) = 0{,}8413$$
$$\Phi(2) = 0{,}9772$$

Por tanto,

$$P(1 \leq Z \leq 2) = \Phi(2) - \Phi(1) = 0{,}9772 - 0{,}8413 = 0{,}1359$$

Gráficamente:



c. En este caso:

$$P(-0,5 \le Z \le -0,2) = P(Z \le -0,2) - P(Z \le -0,5) = \Phi(-0,2) - \Phi(-0,5)$$

En muchos casos las tablas de la normal estándar solo se proporcionan para valores positivos de la variable. Sin embargo, dada la simetría de la distribución normal es fácil verificar que, si $z < 0$,

$$P(Z \le z) = P(Z \ge -z) = 1 - P(Z \le -z)$$

Por tanto, para z < 0 podemos utilizar la relación

$$\Phi(z) = 1 - \Phi(-z)$$

En este caso tendremos:

$$P(-0,5 \le Z \le -0,2) = \Phi(-0,2) - \Phi(-0,5) = (1 - \Phi(0,2)) - (1 - \Phi(-0,5))$$

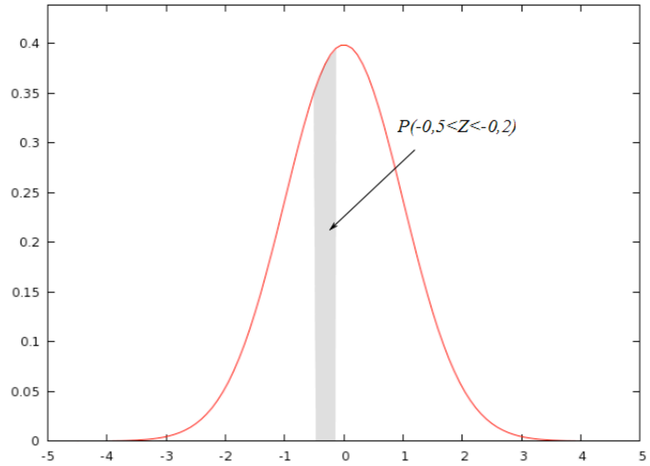
Usando de nuevo las tablas de probabilidad de la normal estándar encontramos:

$$\Phi(0,5) = 0,6915$$
$$\Phi(0,2) = 0,5793$$

Por tanto,

$$P(-0,5 \leq Z \leq -0,2) = (1-\Phi(0,2)) - (1-\Phi(-0,5)) = (1-0,5793) - (1-0,6915)$$
$$P(-0,5 \leq Z \leq -0,2) = 0,4207 - 0,3085 = 0,1122$$

Gráficamente:



d. En este caso tendremos:

$$P(-1 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq -1) = \Phi(2) - \Phi(-1)$$
$$P(-1 \leq Z \leq 2) = \Phi(2) - (1-\Phi(1))$$

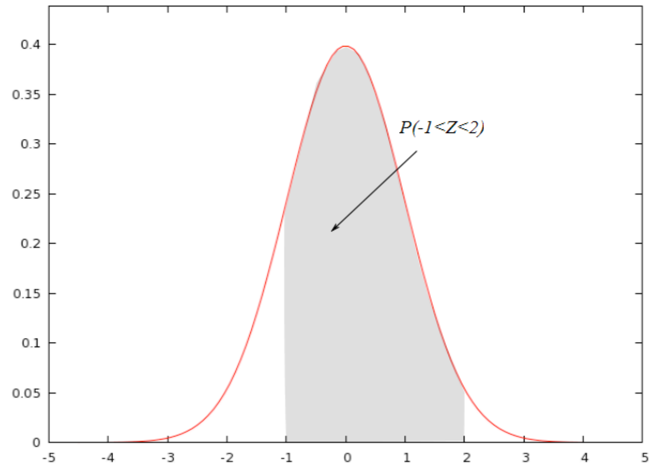
En las tablas de la normal estándar encontramos:

$$\Phi(1) = 0,8413$$
$$\Phi(2) = 0,9772$$

Por tanto,

$$P(-1 \leq Z \leq 2) = \Phi(2) - (1-\Phi(1)) = 0,8413 - (1-0,9772) = 0,8185$$

Gráficamente:



e. Ahora podemos representar la probabilidad buscada como:

$$P(-1 \leq Z \leq \infty) = 1 - P(Z \leq -1) = 1 - \Phi(-1)$$
$$P(-1 \leq Z \leq \infty) = 1 - (1 - \Phi(1)) = \Phi(1)$$

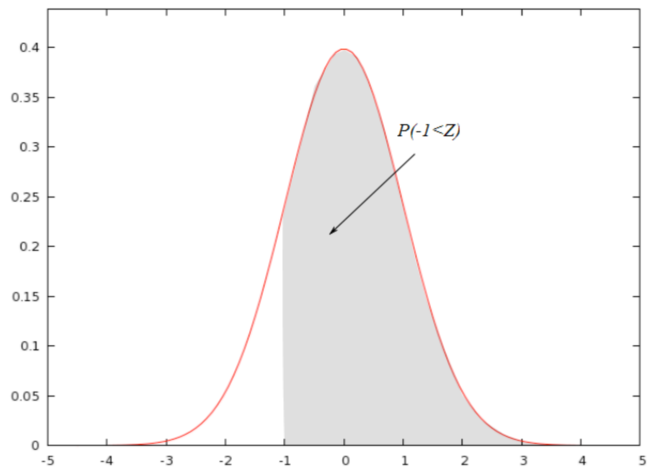
En las tablas de la normal estándar hallamos:

$$\Phi(1) = 0{,}8413$$

Por tanto,

$$P(-1 \leq Z \leq \infty) = \Phi(1) = 0{,}8413$$

Gráficamente:



f. En este apartado la probabilidad buscada viene dada por:

$$P(-\infty \leq Z \leq 1) = P(Z \leq 1) = \Phi(1)$$

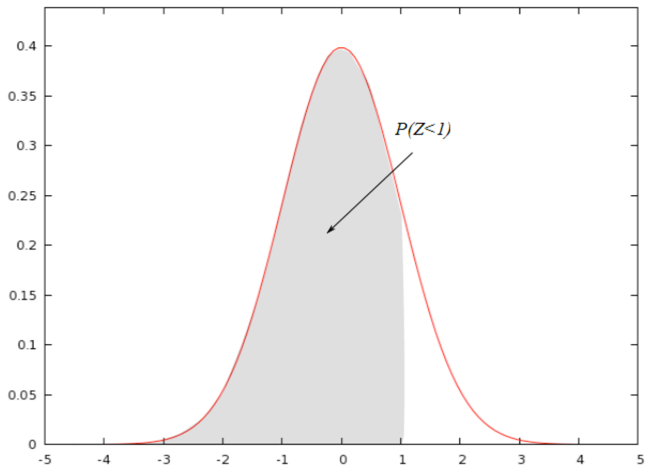
Buscando en las tablas de la normal estándar hallamos:

$$\Phi(1) = 0{,}8413$$

Por tanto,

$$P(-\infty \leq Z \leq 1) = \Phi(1) = 0{,}8413$$

Gráficamente:



Vemos como, por simetría, obtenemos el mismo resultado que en el apartado anterior. Esto es,

$$P(-\infty \leq Z \leq 1) = P(-1 \leq Z \leq \infty)$$

g. Finalmente,

$$P(|Z| \leq 1) = P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1)$$
$$P(|Z| \leq 1) = \Phi(1) - (1 - \Phi(1))$$

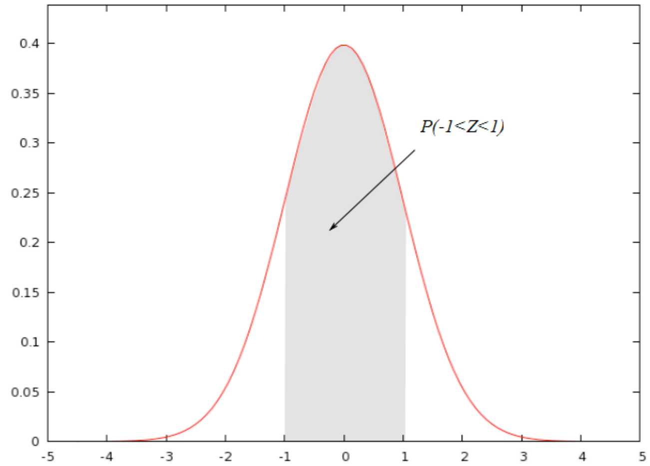
Y una vez más, en las tablas de la normal estándar encontraremos:

$$\Phi(1) = 0{,}8413$$

Por tanto,

$$P(|Z| \leq 1) = \Phi(1) - (1 - \Phi(1)) = 0,8413 - (1 - 0,8413) = 0,6826$$

Gráficamente:



77. Suponer que se sabe que la vida útil de un motor de aviación sigue una distribución de probabilidad de una variable normal con esperanza de 2.000 horas y desviación estándar de 100 horas. Averiguar la probabilidad que un motor dure:

   a. Entre 2.000 y 2.075 horas.
   b. Entre 1.800 y 2.000 horas.
   c. Entre 1.950 y 2.150 horas.
   d. Más de 2.170 horas.
   e. Menos de 1.840 horas.

   a. Sea X la variable que mide la vida útil del motor, de forma que,

$$X \sim N(\ \ ,\sigma) = N(2.000,100)$$

Hemos de hallar $P(2000 \leq X \leq 2075)$ para lo cual debemos estandarizar la variable y utilizar las tablas de la distribución normal estándar:

$$P(2.000 \leq X \leq 2.075) = P(\frac{2.000 -}{\sigma} \leq \frac{X -}{\sigma} \leq \frac{2.075 -}{\sigma}) =$$
$$= P(\frac{2.000 - 2000}{100} \leq Z \leq \frac{2.075 - 2.000}{100}) = P(0 \leq Z \leq 0,75) =$$
$$= P(Z \leq 0,75) - P(Z \leq 0) = \Phi(0,75) - \Phi(0)$$

donde $\Phi(z)$ es el valor correspondiente a $P(Z \leq z)$ que encontramos tabulado. En este caso obtenemos:

$$\Phi(0,75) = 0,7734$$
$$\Phi(0) = 0,5$$

Por tanto,

$$P(2.000 \leq X \leq 2.075) = \Phi(0,75) - \Phi(0) = 0,7734 - 0,5 = 0,2734$$

Existe pues una probabilidad del 27,34% de que el motor dure entre 2.000 y 2.075 horas.

b. Nos piden ahora $P(1.800 \leq X \leq 2.000)$. Debemos de nuevo estandarizar la variable y utilizar las tablas de la distribución normal estándar:

$$P(1.800 \leq X \leq 2.000) = P(\frac{1.800 -}{\sigma} \leq \frac{X -}{\sigma} \leq \frac{2.000 -}{\sigma}) =$$
$$= P(\frac{1.800 - 2.000}{100} \leq Z \leq \frac{2.000 - 2.000}{100}) = P(-2 \leq Z \leq 0) =$$
$$= P(Z \leq 0) - P(Z \leq -2) = \Phi(0) - \Phi(-2) = \Phi(0) - (1 - \Phi(2))$$

donde $\Phi(-2) = P(Z \leq -2) = P(Z \geq 2) = (1 - P(Z \leq 2) = (1 - \Phi(2))$ por la simetría de la distribución normal. Consultando las tablas de la normal estándar obtenemos:

$$\Phi(2) = 0,9772$$
$$\Phi(0) = 0,5$$

Por tanto,

$$P(1.800 \le X \le 2.000) = \Phi(0) - (1 - \Phi(2)) = 0{,}5 - 0.0228 = 0{,}4772$$

Existe pues una probabilidad del 47,72% de que el motor dure entre 1.800 y 2.000 horas.

c. Procediendo de forma similar tenemos ahora que

$$P(1.950 \le X \le 2.150) = P(\frac{1.950 -}{\sigma} \le \frac{X -}{\sigma} \le \frac{2.150 -}{\sigma}) =$$
$$= P(\frac{1.950 - 2.000}{100} \le Z \le \frac{2.150 - 2.000}{100}) = P(-0{,}5 \le Z \le 1{,}5) =$$
$$= P(Z \le 1{,}5) - P(Z \le -0{,}5) = \Phi(1{,}5) - \Phi(-0{,}5) = \Phi(1{,}5) - (1 - \Phi(0{,}5))$$

Donde de nuevo $\Phi(-0{,}5) = (1 - \Phi(0{,}5))$ por la simetría de la distribución normal. Consultando las tablas de la normal estándar obtenemos:

$$\Phi(0{,}5) = 0{,}6915$$
$$\Phi(1{,}5) = 0{,}9332$$

Por tanto,

$$P(1.950 \le X \le 2.150) = \Phi(1{,}5) - (1 - \Phi(0{,}5)) = 0{,}9332 - 0.3085 = 0{,}6247$$

Existe pues una probabilidad del 62,47% de que el motor dure entre 1.950 y 2.150 horas.

d. En este caso,

$$P(X \ge 2.170) = 1 - P(X \le 2.170) =$$
$$= 1 - P(\frac{X -}{\sigma} \le \frac{2170 - 2000}{100}) = 1 - P(Z \le 1{,}7) = 1 - \Phi(1{,}7)$$

Consultando las tablas de la normal estándar obtenemos:

$$\Phi(1,7) = 0,9554$$

Por tanto,

$$P(X \geq 2.170) = 1 - \Phi(1,7) = 1 - 0,9554 = 0,0446$$

Existe pues una probabilidad del 4,46% de que el motor dure más de 2.170 horas.

e. Finalmente,

$$P(X \leq 1.840) = P(\frac{X -}{\sigma} \leq \frac{1.840 - 2.000}{100}) = P(Z \leq -1,6) = \Phi(-1,6) = 1 - \Phi(1,6)$$

Consultando las tablas de la normal estándar obtenemos:

$$\Phi(1,6) = 0,9452$$

Por tanto,

$$P(X \geq 1.840) = 1 - \Phi(-1,6) = 1 - 0,9452 = 0,0548$$

Existe pues una probabilidad del 5,48% de que el motor dure menos de 1.840 horas.

78. (*) Suponer que las notas de un curso de Estadística se distribuyen normalmente con media de 5,5 y desviación estándar de 1.

a. Determinar la proporción de alumnos que han obtenido entre 6 y 7 puntos.
b. Determinar la proporción de suspensos en la clase.
c. Determinar la nota que delimita al 10% de los alumnos con notas más altas.

a. Si la variable X recoge la nota del curso de Estadística tenemos que, según la información proporcionada, $X \sim N(\ \ ,\sigma) = N(5{,}5;1)$
Por tanto, para hallar la proporción de alumnos con notas comprendidas entre 6 y 7, es decir, $P(6 \le X \le 7)$, deberemos estandarizar la variable para poder utilizar los valores tabulados.

$$P(6 \le X \le 7) = P(\frac{6-}{\sigma} \le \frac{X-}{\sigma} \le \frac{7-}{\sigma}) =$$
$$P(\frac{6-5{,}5}{1} \le Z \le \frac{7-5{,}5}{1}) = P(0{,}5 \le Z \le 1{,}5) =$$
$$P(Z \le 1{,}5) - P(Z \le 0{,}5) = \Phi(1{,}5) - \Phi(0{,}5)$$

dónde $\Phi(z)$ es el valor correspondiente a $P(Z \le z)$ que encontramos en las tablas de la normal estándar $N(0,1)$. Consultando estas tablas obtenemos:

$$\Phi(0{,}5) = 0{,}6915$$
$$\Phi(1{,}5) = 0{,}9332$$

Por tanto,

$$P(6 \le X \le 7) = \Phi(1{,}5) - \Phi(0{,}5) = 0{,}9332 - 0{,}6915 = 0{,}2417$$

Es decir, un 24,17% de los alumnos obtuvo una nota entre 6 y 7.

b. La proporción de alumnos que ha suspendido la asignatura, es decir, que ha obtenido una nota inferior a 5, $P(X \le 5)$, la obtenemos de forma similar al apartado anterior.

$$P(X \le 5) = P(\frac{X-}{\sigma} \le \frac{5-}{\sigma}) = P(Z \le \frac{5-5{,}5}{1}) =$$
$$= P(Z \le -0{,}5) = \Phi(-0{,}5) = 1 - \Phi(0{,}5)$$

donde $\Phi(-0{,}5) = 1 - \Phi(0{,}5)$ por la simetría de la normal estándar $N(0{,}1)$.
Consultando estas tablas obtenemos:

$$\Phi(0{,}5) = 0{,}6915$$

Por tanto,

$$P(X \leq 5) = 1 - \Phi(0{,}5) = 1 - 0{,}6915 = 0{,}3085$$

Es decir, un 30,85% de los alumnos suspendió la asignatura.

c. Nos están pidiendo hallar la nota k tal que el 10% de los alumnos han obtenido una nota superior a k. Es decir, hallar k tal que se verifica:

$$P(X \geq k) = 0{,}1$$

Estandarizando la variable y procediendo como en los apartados anteriores obtenemos:

$$P(X \geq k) = 0{,}1 \Rightarrow P(\frac{X-}{\sigma} \geq \frac{k-}{\sigma}) = 0{,}1 \Rightarrow$$
$$\Rightarrow P(Z \geq \frac{k-5{,}5}{1}) = 0{,}1 \Rightarrow 1 - P(Z \leq k-5{,}5) = 0{,}1 \Rightarrow$$
$$\Rightarrow P(Z \leq k-5{,}5) = 0{,}9$$

En este caso, al utilizar las tablas, conocemos el valor de la probabilidad (0,9) y debemos buscar el valor de la variable que más se aproxime a esta probabilidad. Es decir, hallar el valor z en la tabla de la normal estándar tal que $\Phi(z) = 0{,}9$ (o se aproxime lo máximo posible a 0,9). Consultando las tablas observamos que

$$\Phi(1{,}28) = 0{,}8997$$
$$\Phi(1{,}29) = 0{,}9015$$

Por tanto, tomamos el valor z = 1,28 como el más próximo:

$$\Phi(1{,}28) \approx 0{,}9 \Rightarrow P(Z \leq 1{,}28) \approx 0{,}9$$

Recapitulando, después de estandarizar estábamos buscando el valor *k* tal que:

$$P(Z \leq k - 5{,}5) = 0{,}9$$

Y hemos obtenido que:

$$P(Z \leq 1{,}28) = 0{,}9$$

Por tanto,

$$k - 5{,}5 = 1{,}28 \Rightarrow k = 5{,}5 + 1{,}28 = 6{,}78$$

En conclusión, el 10% de los alumnos han obtenido una nota superior al 6,78.

79. (*) Dos planes de inversión A y B están caracterizados por haber obtenido una rentabilidad media, en porcentaje, de μ_A= 10.4 y μ_B= 11, respectivamente. La dispersión de la rentabilidad de ambos planes, medida por la varianza, ha sido de $\sigma_A^2 = 1{,}44$ y $\sigma_B^2 = 17{,}64$, respectivamente. Determinar qué plan de inversión ofrece mayor probabilidad de obtener una rentabilidad de por lo menos un 10% (asumir una distribución normal para las rentabilidades).

Haremos que X_A y X_B representen las dos variables normales que describen las rentabilidades de ambos planes. Necesitamos calcular:

$$P(X_A \geq 10)$$
$$P(X_B \geq 10)$$

y ver cuál de ellas es mayor. Usamos la normal estándar:

$$P(X_A \geq 10) = P\ \frac{X_A - \ _A}{\sigma_A} \geq \frac{10 - 10{,}4}{1{,}2} \ = P(Z \geq -0{,}3333)$$
$$P(X_B \geq 10) = P\ \frac{X_B - \ _B}{\sigma_B} \geq \frac{10 - 11}{4{,}2} \ = P(Z \geq -0{,}2381)$$

Usando las tablas de la probabilidad normal encontramos que $P(X_A \geq 10)=0{,}6293$ y $P(X_B \geq 10)=0{,}5910$. A pesar que el plan *A* ofrece una rentabilidad promedio inferior, el inversor, en términos de la probabilidad debería optar por *A*, en lugar de *B*.

80. Se sabe que en un taller de coches la sustitución de una pieza defectuosa requiere un tiempo medio de 75 minutos con una desviación estándar de 20 minutos.

   a. ¿Qué proporción de sustituciones se hacen en menos de una hora?
   b. ¿Qué proporción en más de noventa minutos?
   c. 0.1 es la probabilidad que una sustitución se haga en más de, ¿cuántos minutos?

   a. Convertimos en primer lugar las horas y minutos a formato decimal, de forma que 75 minutos son 1,25 horas y 20 minutos son 1/3 horas.
   A partir de aquí, si X es la variable que mide el tiempo que tarda la sustitución de una pieza defectuosa tenemos que $X \sim N(\;\;,\sigma) = N(1{,}25;\ 1/3)$. En este caso nos están pidiendo la proporción de sustituciones que se hacen en menos de una hora, es decir, $P(X \leq 1)$. Estandarizando la variable X tenemos

$$P(X \leq 1) = P\left(\frac{X -\;}{\sigma} \leq \frac{1 -\;}{\sigma}\right) = P\left(Z \leq \frac{1 - 1{,}25}{1/3}\right) = P(Z \leq -0{,}75) = \Phi(-0{,}75)$$

   Aprovechando la simetría de la distribución normal sabemos que:

$$\Phi(-0{,}75) = 1 - \Phi(0{,}75)$$

   Consultando las tablas de la normal estándar vemos que:

$$\Phi(0{,}75) = 0{,}7734$$

   Por tanto, la proporción que se pide es:

$$P(X \leq 1) = \Phi(-0{,}75) = 1 - \Phi(0{,}75) = 1 - 0{,}7734 = 0{,}2266$$

   Es decir, un 22,66% de las sustituciones se realizarán en menos de 1 hora.

   b. Nos piden ahora la proporción de sustituciones que se hacen en más de una hora y media, es decir, $P(X \geq 1{,}5)$. Estandarizando la variable X tenemos:

$$P(X \geq 1{,}5) = 1 - P(X \leq 1{,}5) = 1 - P\left(\frac{X -\;}{\sigma} \leq \frac{1{,}5 -\;}{\sigma}\right) =$$
$$= 1 - P\left(Z \leq \frac{1{,}5 - 1{,}25}{(1/3)}\right) = 1 - P(Z \leq 0{,}75) = 1 - \Phi(0{,}75)$$

En el apartado anterior habíamos encontrado que:

$$\Phi(0,75) = 0,7734$$

Por tanto, la proporción que se pide ahora es:

$$P(X \geq 1,5) = 1 - \Phi(0,75) = 1 - 0,7734 = 0,2266$$

Es decir, un 22,66% de las sustituciones se realizarán en más de 90 minutos.

c. Nos piden ahora hallar el tiempo k tal que con una probabilidad de 0,1 la sustitución se haga en más de k horas. Es decir, hemos de hallar k tal que

$$P(X \geq k) = 0,1$$

Estandarizando tenemos,

$$P(X \geq k) = 0,1 \Rightarrow P(\frac{X -}{\sigma} \geq \frac{k -}{\sigma}) = 0,1 \Rightarrow$$
$$\Rightarrow P(Z \geq \frac{k - 1,25}{1/3}) = 0,1 \Rightarrow P(Z \geq 3k - 3,75) = 0,1 \Rightarrow$$
$$\Rightarrow 1 - P(Z \leq 3k - 3,75) = 0,1 \Rightarrow P(Z \leq 3k - 3,75) = 0,9$$

Es decir, en las tablas de la normal estándar hemos de hallar el valor z tal que $\Phi(z) = P(Z \leq z) = 0,9$ (o se aproxime lo más posible). Así observamos que:

$$\Phi(1,28) = 0,8997$$
$$\Phi(1,29) = 0,9015$$

y escogemos $z = 1,28$ como el valor cuya probabilidad está más cercana a 0,9.
Por tanto, por una parte teníamos que hallar k tal que $P(Z \leq 3k - 3,75) = 0,9$, y por otra hemos visto que $P(Z \leq 1,28) = 0,9$. Combinando las dos expresiones tenemos que:

$$3k-3{,}75=1{,}28 \Rightarrow 3k=1{,}28+3{,}75 \Rightarrow$$
$$\Rightarrow 3k=5{,}03 \Rightarrow k=\frac{5{,}03}{3} \Rightarrow k=1{,}677 \approx 1:40$$

En definitiva, con una probabilidad del 10% la sustitución se realizará en más de 1 hora y 40 minutos.

81. (*) Una empresa produce arandelas. La experiencia demuestra que la media de los diámetros de las arandelas es 1 cm y la desviación estándar es de 0,01 cm. El uso de las arandelas permite una tolerancia máxima en el diámetro de 0,98 a 1,02 cm. En caso contrario las arandelas se consideran defectuosas. Determinar el porcentaje de arandelas defectuosas que produce la máquina (asumir que los diámetros se distribuyen normalmente).

El supuesto de normalidad nos indica que si X es la variable que mide el diámetro de las arandelas entonces $X \sim N(\ \ ,\sigma)=N(1,\ 0{,}01)$.
Las arandelas “correctas” son aquellas cuyo diámetro está entre 0,98 y 1,02 cm. Por tanto, para hallar la proporción de arandelas “correctas” hemos de calcular:

$$P(0{,}98 \le X \le 1{,}02)=P(\frac{0{,}98-\ }{\sigma} \le \frac{X-\ }{\sigma} \le \frac{1{,}02-\ }{\sigma})=$$
$$=P(\frac{0{,}98-1}{0{,}01} \le Z \le \frac{1{,}02-1}{0{,}01})=P(-2 \le Z \le 2)=$$
$$=P(Z \le 2)-P(Z \le -2)=\Phi(2)-\Phi(-2)=\Phi(2)-(1-\Phi(2)))$$

donde la última parte de la igualdad se obtiene por la simetría de la distribución normal de forma que $\Phi(-2)=1-\Phi(2)$. Buscamos este valor en las tablas y obtenemos $\Phi(2)=0{,}9772$. Por tanto, la proporción de arandelas “correctas” que estábamos buscando es:

$$P(0{,}98 \le X \le 1{,}02)=\Phi(2)-(1-\Phi(2)))=0{,}9772-(1-0{,}9772)=0{,}9544$$

Finalmente, si hemos obtenido que una proporción del 95,44% son arandelas “correctas”, la proporción de arandelas defectuosas (que es lo que originalmente se nos pide) será de:

$$1-0{,}9544=0{,}0456$$

En definitiva, un 4,56% de las arandelas resultarán defectuosas.

82. (*) Vamos a asumir que los ingresos familiares anuales siguen las pautas de una variable aleatoria normal. Se ha evaluado, además, que la desviación estándar de esta variable normal es σ = 20.000 (euros). Hacienda informa que el 4% de las unidades familiares declaran ingresos superiores a 60.000 euros anuales.

   a. Calcular cuál es el ingreso anual promedio de una unidad familiar.
   b. Evaluar qué proporción de familias tienen ingresos anuales inferiores a 12.000 euros (los denominados “mileuristas”).
   c. Evaluar también que proporción de familias tienen ingresos anuales entre 20.000 y 40.000 euros (la clase media).

a. Tenemos una variable aleatoria normal $X \sim N(\ ,\sigma) = N(\ ,20.000)$ de la que desconocemos el valor de μ y necesitamos obtenerlo. Sabemos lo siguiente:

$$P(X \geq 60.000) = 0,04$$

Expresaremos esta probabilidad conocida en términos de la variable normal estándar:

$$P(X \geq 60.000) = P \ \frac{X -}{\sigma} \geq \frac{60.000 -}{20.000} \ = 0,04$$

En las tablas de probabilidad de la normal estándar hemos de hallar el valor de z^* tal que $P(Z \geq z^*) = 0,04$. Este valor resulta ser z^*=1,75. De aquí:

$$\frac{60.000 -}{20.000} = 1,75$$

de donde se sigue que μ=25.000

b. Estamos ya en condiciones de evaluar $P(X \leq 12.000)$. Pasamos a la normal estándar y hallamos:

$$P(X \leq 12.000) = P \ \frac{X -}{\sigma} \leq \frac{12.000 - 25.000}{20.000} \ = P(Z \leq -0,65) = 0,2578$$

c. Ahora queremos evaluar $P(20.000 \leq X \leq 40.000)$. Procedemos de igual manera pasando a la normal estándar y obteniendo la probabilidad buscada de la tabla:

$$P(20.000 \leq X \leq 40.000) = P \quad \frac{20.000 - 25.000}{20.000} \leq Z \leq \frac{40.000 - 25.000}{20.000} \quad =$$

$$= P(-0,25 \leq Z \leq 0,75) = 0,3721$$

83. Una empresa dispone de una máquina que llena botellas de cola. Se sabe que el llenado de las botellas sigue las pautas de una variable normal con media μ= 34 (cl.) y varianza σ^2 = 2,25. Las botellas que contienen menos de 33 cl se retiran de la circulación.

   a. Determinar el porcentaje de botellas que se retiran de la circulación comercial.
   b. Si seleccionamos 10 botellas al azar, determinar la probabilidad que ninguna tenga que ser retirada de la circulación.
   c. La máquina permite modular el volumen promedio de líquido. ¿Cuál sería el volumen medio para que solo el 1,5% de las botellas contengan menos de 33 cl?

   a. Para esta variable la desviación estándar es de σ = 1,5. El porcentaje de botellas desechadas se corresponde con la proporción de botellas cuyo volumen de líquido está por debajo de 33 cl. En otras palabras, necesitamos evaluar $P(X<33)$. Transformamos en primer lugar la variable normal X en una variable normal estándar Z y usando la tabla de probabilidad hallamos el porcentaje buscado:

$$P(X < 33) = P \quad Z < \frac{33 -}{\sigma} \quad = P \quad Z < \frac{33 - 34}{1,5} \quad = P(Z < -0,66) = 0,2546$$

   b. Cada una de las 10 botellas puede ser desechada (con probabilidad 25,46%) o no (con probabilidad 74,54%). Esto se corresponde con una variable binomial con n = 10 y p = 0,2546. Deseamos evaluar la probabilidad que no se deseche ninguna botella, es decir, que X = 0:

$$P(X = 0) = C(10,0) \cdot p^0 \cdot (1 - p)^0 = 0,05295 \approx 5,3\%$$

   c. Se trata ahora de determinar el nuevo promedio μ a partir del cual solo un 1,5% de las botellas serían desechadas al no satisfacer el requisito de los 33 cl de volumen. Partimos pues de saber que debe cumplirse que $P(X<33)$ = 0,015. Pasando a la normal estándar:

$$P(X < 33) = P \quad Z < \frac{33 -}{1,5} \quad = 0,015$$

Necesitamos especificar el valor z tal que hasta él se acumula un área de 0,015 por debajo de la densidad normal. Usando la tabla de probabilidad normal concluimos que el valor de z asociado a esta probabilidad es z = -2,17. En consecuencia:

$$z = -2{,}17 = \frac{33 -}{1{,}5}$$

y de aquí resolviendo para μ se sigue que μ = 36,26.

84. (*) Después de realizar un examen de Estadística I vemos que las notas se han distribuido siguiendo las pautas de una variable normal. El profesor ha calculado la nota media y la varianza de las notas, que han resultado ser $\quad = 5{,}5$ y $\sigma^2 = 2{,}25$, respectivamente. Las notas satisfacen la escala tradicional de puntuaciones de 0 a 10. Con esta información se pide:

   a. Si se escoge un estudiante al azar, evaluar la probabilidad que su nota sea menor o igual a 2.
   b. Si se escogen 5 estudiantes al azar, evaluar la probabilidad que uno de ellos haya obtenido una nota menor o igual a 2.
   c. Si en total hay 1.000 estudiantes de Estadística I, estimar la probabilidad que entre estos 1.000 hayan 15 estudiantes con nota menor o igual a 2.

   a. Sea X la normal que gobierna las notas. Hemos de evaluar $P(X \leq 2)$. Pasaremos a la normal estándar y buscaremos la probabilidad en las tablas de probabilidad:

$$P(X \leq 2) = P(Z \leq \frac{2-\mu}{\sigma}) = P(Z \leq \frac{2-5{,}5}{1{,}5}) = P(Z \leq -2{,}33) = 0{,}0099$$

   b. Obtener una nota menor o igual que 2 se puede considerar un buen "fracaso". Hemos evaluado su probabilidad en 0,0099. Que haya 1 estudiante entre 5 con este "fracaso" es un suceso cuya probabilidad responde a la ley binomial. Hagamos que Y represente una variable binomial con n = 5 y p = 0,0099. La probabilidad buscada se puede calcular como:

$$P(Y = 1) = C(5{,}1) \cdot p^1 \cdot (1-p)^4 = 0{,}0476 = 4{,}76\%$$

   c. Volvemos a tener una situación binomial pero ahora con n = 1.000. Tendríamos que calcular:

$$P(Y = 15) = C(1.000{,}15) \cdot p^{15} \cdot (1-p)^{(1.000-15)}$$

Ahora bien, es más simple usar la aproximación de Poisson a la binomial tomando:

$$\lambda = n \cdot p = 1.000 \cdot 0{,}0099 = 9{,}9$$

De aquí:

$$P(Y = 15) = \frac{e^{-\lambda} \cdot \lambda^{15}}{15!} = 0{,}033 = 3{,}3\%$$

85. Volviendo al examen de Estadística I de la pregunta anterior se desea saber:

   a. El porcentaje de estudiantes que han superado la materia (se supera con un mínimo de 5).
   b. El porcentaje que han obtenido como mínimo un notable (nota que se obtiene sacando un mínimo de 8).
   c. La nota de corte que delimita al 10% de los estudiantes con mejor nota.

   a. Sea de nuevo X la variable normal que gobierna la distribución de las notas. Hemos de evaluar $P(X \geq 5)$. Lo hacemos pasando a la normal estándar y usando las tablas de probabilidad:

$$P(X \geq 5) = P(Z \geq \frac{5-}{\sigma}) = P(Z \geq \frac{5-5{,}5}{1{,}5}) = P(Z \geq -0{,}333) = 0{,}6293$$

   b. Aquí se procede como en el apartado anterior pero ahora para $P(X \geq 8)$:

$$P(X \geq 8) = P(Z \geq \frac{8-}{\sigma}) = P(Z \geq \frac{8-5{,}5}{1{,}5}) = P(Z \geq 1{,}6667) = 0{,}0475$$

   c. En esta situación hemos de evaluar el valor de z, que para la normal estándar produce un valor de probabilidad en la cola alta igual al 10%: $P(Z \geq z) = 0{,}10$. Usando la tabla de la probabilidad normal averiguamos que el valor de z estará entre 1,28 y 1,29. Así, el área entre $z = 0$ y $z = 1{,}28$ es prácticamente 0,40. Aproximaremos el resultado tomando pues $z = 1{,}28$. El valor de la normal no-estándar y que da lugar a $z = 1{,}28$ lo podemos obtener a partir de:

$$x = \sigma \cdot z + \mu = 1{,}5 \cdot 1{,}28 + 5{,}5 = 7{,}42$$

La nota de corte para el 10% superior es pues de 7,42 en la escala de 0 a 10.

86. El tiempo dedicado a un estudiante que acude a las tutorías de un profesor es en promedio de diez minutos. Suponer que el tiempo dedicado a un estudiante sigue una distribución exponencial.

   a. Averiguar la probabilidad que un estudiante esté menos de veinte minutos con el profesor.
   b. Ídem que esté más de cinco minutos.
   c. Ídem que esté entre cinco y veinte minutos.

   a. El tiempo medio dedicado a un estudiante es de 10 minutos; en otras palabras, entre dos ocurrencias (visitas de estudiantes) transcurren 10 minutos, en promedio. Si el proceso sigue una variable exponencial, la función de densidad de esta variable aleatoria X está dada por:

$$f_X(x) = \frac{e^{-\frac{x}{\ }}}{\ }$$

siendo $x \geq 0$ y con $\ = 10$ en este caso concreto. Usando métodos de cálculo integral es posible ver que la función de probabilidad acumulada tiene la siguiente estructura:

$$F_X(x) = P(0 \leq X \leq x) = 1 - e^{-\frac{x}{\ }}$$

En este apartado, y para $\ = 10$, queremos evaluar cuál es la probabilidad que un alumno esté menos de 20 minutos con el profesor:

$$F_X(20) = P(0 \leq X \leq 20) = 1 - e^{-\frac{20}{\ }} = 0{,}8647 = 86{,}47\%$$

   b. Estar más de 5 minutos con el profesor corresponde a:

$$P(X > 5) = 1 - P(0 \leq X \leq 5) = 1 - \ \ 1 - e^{-\frac{5}{\ }} \ \ = 0{,}6065 = 60{,}65\%$$

c. Estar entre 5 y 20 minutos siendo atendido por el profesor tiene probabilidad evaluable por:

$$P(5 \leq X \leq 20) = P(0 \leq X \leq 20) - P(0 \leq X \leq 5) =$$

$$1-e^{-\frac{20}{}} \quad - \quad 1-e^{-\frac{5}{}} \quad = 0,4712 = 47,12\%$$

87. La red local de ordenadores de la facultad tiene en promedio una caída cada dos semanas. Suponga que la red acaba de fallar. Averigüe la probabilidad de que pase un mes antes de la próxima caída del sistema.

En el periodo en cuestión el sistema cae dos veces. Esto corresponde con un fenómeno de Poisson con parámetro λ = 2. El tiempo promedio entre dos caídas es por consiguiente de 1/2=0,5 meses y podemos analizar el fenómeno de las caídas usando una variable exponencial X con parámetro $\quad = 0,5$. Para este valor, la probabilidad que el sistema de ordenadores no caiga en el próximo mes puede evaluarse por:

$$P(X > 1) = 1 - P(0 \leq X \leq 1) = 1 - e^{-\frac{1}{}} = 1 - 0,8647 = 0,1353 = 13,53\%$$

88. En los últimos años una empresa ha sufrido un promedio de dos huelgas al año. La semana pasada concluyó la huelga realizada para presionar en una negociación. Averiguar la probabilidad de que:

a. Transcurra un año sin otra huelga.
b. Transcurra medio año sin otra huelga.
c. Transcurran dos años sin otra huelga.

a. Igual que en el ejercicio anterior, vemos que el número de ocurrencias (huelgas) en el periodo considerado es de 2. Para este fenómeno de ocurrencias infrecuentes en el intervalo temporal de un año tendremos una situación modelizable con una variable de Poisson con λ = 2, de manera que el tiempo promedio entre 2 huelgas será de $\quad = 1/\lambda = 1/2 = 0,5$. Que transcurra un año sin huelgas tiene probabilidad evaluable por una variable exponencial, de manera que:

$$P(X > 1) = 1 - P(0 \leq X \leq 1) = 1 - e^{-\frac{1}{}} = 1 - 0,8647 = 0,1353 = 13,53\%$$

b. Para esta segunda posibilidad, habremos de evaluar:

$$P(X > 0{,}5) = 1 - P(0 \leq X \leq 0{,}5) = 1 - e^{-\frac{1}{\;}} = 1 - 0{,}6321 = 0{,}3679 = 36{,}79\%$$

c. Finalmente, ahora deberíamos calcular:

$$P(X > 2) = 1 - P(0 \leq X \leq 2) = 1 - e^{-\frac{1}{\;}} = 1 - 0{,}9817 = 0{,}0183 = 1{,}83\%$$

En definitiva, la probabilidad que no haya una huelga en los próximos dos años es extremadamente baja.

Vectores aleatorios

89. Consideramos dos experimentos de azar que utilizan monedas idénticas y honestas. En el primero lanzamos 2 monedas al aire y medimos el número de caras con una variable aleatoria X; en el segundo lanzamos una única moneda y medimos las caras con una variable Y. A continuación construimos la variable suma $Z = X + Y$.

    a. ¿Qué propiedad se puede establecer entre la esperanza matemática de Z y las esperanzas de X y de Y?
    b. Verificar si esa propiedad se cumple efectuando todos los cálculos pertinentes y justificando todos los pasos realizados.

    a. Puesto que $Z = X + Y$ la propiedad en cuestión es que la esperanza de una suma es la suma de las esperanzas: $E(Z) = E(X + Y) = E(X) + E(Y)$.

    b. X toma valores 0, 1 y 2 con probabilidades 0,25, 0,5 y 0,25, respectivamente. La variable Y toma valores 0 y 1 con probabilidades respectivas de 0,5. La variable suma Z tomará en consecuencia valores 0, 1, 2 y 3. Se trata ahora de especificar la distribución de probabilidad de Z.

    En primer lugar, (Z = 0) si y solo si (X = 0 y Y= 0). Puesto que los dos experimentos son independientes podemos usar la regla multiplicativa de la probabilidad que nos permite separar las probabilidades absolutas:

$$P(Z=0) = P(X=0 \text{ y } Y=0) = P(X=0)\cdot P(Y=0) = (0{,}25)\cdot(0{,}5) = 0{,}125$$

    Tendremos (Z = 1) si y solo si (X = 0 y Y = 1) o (X = 1 y Y = 0). Estos dos sucesos son excluyentes y de ahí que podamos usar la regla aditiva de la probabilidad a los sucesos excluyentes y la regla multiplicativa a cada uno de los sucesos integrantes:

$$\begin{aligned} P(Z=1) &= P\left[(X=0) \text{ y } (Y=1)\right] + P\left[(X=1) \text{ y } (Y=0)\right] = \\ &= (P(X=0)\cdot P(Y=1)) + (P(X=1)\cdot P(Y=0)) \\ &= (0{,}25)\cdot(0{,}5) + (0{,}5)\cdot(0{,}5) = 0{,}375 \end{aligned}$$

    Tendremos (Z = 2) si y solo si (X = 1 y Y = 1) o (X = 2 y Y = 0). En consecuencia y por los mismos argumentos anteriores:

$$\begin{aligned} P(Z=2) &= P\left[(X=1) \text{ y } (Y=1)\right] + P\left[(X=2) \text{ y } (Y=0)\right] = \\ &= (P(X=1)\cdot P(Y=1)) + (P(X=2)\cdot P(Y=0)) \\ &= (0{,}5)\cdot(0{,}5) + (0{,}25)\cdot(0{,}5) = 0{,}375 \end{aligned}$$

Finalmente, ($Z = 3$) si y solo si ($X = 2$ y $Y = 1$). En este caso:

$$P(Z=3)=P\left[(X=2)\text{ y }(Y=1)\right]=P(X=2)\cdot P(Y=1)=(0{,}25)\cdot(0{,}5)=0{,}125$$

Podemos verificar que las probabilidades asociadas a Z suman en total 1. Calculamos las esperanzas y comprobamos que la propiedad comentada en *a*) se cumple:

$$E(X)=0\cdot P(X=0)+1\cdot P(X=1)+2\cdot P(X=2)=1$$
$$E(Y)=0\cdot P(Y=0)+1\cdot P(Y=1)=0{,}5$$
$$E(Z)=0\cdot P(Z=0)+1\cdot P(Z=1)+2\cdot P(Z=2)+3\cdot P(Z=3)=1{,}5$$

90. Considere un par de variables aleatorias (X, Y) que toman los siguientes valores: (1,0), (0,1), (– 1,0) y (0, – 1), todos ellos con la misma probabilidad.

a. Construir la matriz de probabilidades del par (X, Y) (es decir, la distribución de probabilidad conjunta en formato tabular).
b. Calcular las funciones de probabilidad marginal.
c. Calcular $E(X)$ y $E(Y)$.
d. Considerar la nueva variable aleatoria $X\cdot Y$ (el producto de X e Y). Hallar $E(X\cdot Y)$ a partir de la distribución de probabilidad de $X\cdot Y$.
e. Verificar si $E(X\cdot Y) = E(X)\cdot E(Y)$.
f. Discutir si las variables aleatorias X e Y son independientes.
g. Calcular la covarianza cov(X, Y) del par aleatorio (X, Y). Discutir los resultados que obtenga en los apartados *e* y *f*.

a. Si todos los 4 valores indicados son equiprobables (p=1/4) tendremos la siguiente tabla de probabilidad para la variable bidimensional (X, Y):

X / Y	*1*	*0*	*-1*
1	0	1/4	0
0	1/4	0	1/4
-1	0	1/4	0

b. Obtenemos las funciones de probabilidad marginales de X (sumando por filas) e Y (sumando por columnas):

X / Y	*1*	*0*	*-1*	f_X
1	0	1/4	0	1/4
0	1/4	0	1/4	2/4
-1	0	1/4	0	1/4
f_Y	1/4	2/4	1/4	

c. Dado que,

$$E(X) = \sum_{x} x \cdot f_X(x)$$

Tendremos:

$$E(X) = 1 \cdot \frac{1}{4} + 0 \cdot \frac{2}{4} + (-1) \cdot \frac{1}{4} = 0$$
$$E(Y) = 1 \cdot \frac{1}{4} + 0 \cdot \frac{2}{4} + (-1) \cdot \frac{1}{4} = 0$$

d. Evaluamos la esperanza del producto:

$$E(X \cdot Y) = \sum_{(x,y)} (x \cdot y) \cdot f_{XY}(x,y)$$

que en este caso da lugar a:

$$E(X \cdot Y) = (1 \cdot 1) \cdot 0 + (1 \cdot 0) \cdot \frac{1}{4} + (1 \cdot (-1)) \cdot 0 + (0 \cdot 1) \cdot \frac{1}{4}$$
$$+ (0 \cdot 0) \cdot 0 + (0 \cdot (-1)) \cdot \frac{1}{4} + ((-1) \cdot 1) \cdot 0 + ((-1) \cdot 0) \cdot \frac{1}{4} + ((-1) \cdot (-1)) \cdot 0 = 0$$

e. Efectivamente, en este caso:

$$E(X \cdot Y) = 0 = 0 \cdot 0 = E(X) \cdot E(Y)$$

f. Claramente NO son independientes. Para que lo fueran sería necesario que

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

y sin embargo comprobamos que

$$f_{XY}(1,1) = 0 \neq \frac{1}{4} \cdot \frac{1}{4} = f_X(1) \cdot f_Y(1)$$

g. Sabemos que

$$\operatorname{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Además hemos visto que

$$E(X \cdot Y) = E(X) \cdot E(Y) = 0$$

y de aquí que cov(X,Y) = 0. Sin embargo, las variables NO son independientes aunque resulte que la covarianza entre ellas sea nula. Este ejemplo ilustra el hecho que si dos variables aleatorias son independientes, entonces su covarianza es cero, mientras que la afirmación contraria no es cierta en general: covarianza cero no implica necesariamente que las variables sean independientes (como acabamos de ver en el ejemplo).

91. Tenemos un par de variables aleatorias X e Y que toman valores X= –1, 0, +1, y Y =0, +1. La tabla de probabilidades conjuntas está dada por:

X/Y	*0*	*+1*
–1	0	0.25
0	0.5	0
+1	0	0.25

a. Calcular la covarianza entre X e Y.
b. Discernir si X e Y son, o no son, variables aleatorias independientes y explicar en su caso el porqué.

a. Una vez más la covarianza se puede calcular a partir de:

$$\operatorname{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Para este cálculo primero necesitamos las distribuciones marginales de probabilidad de X y de Y. Una vez obtenidas (sumando las filas para X y las columnas para Y) encontraríamos:

$$E(X) = -1 \cdot 0{,}25 + 0 \cdot 0{,}5 + 1 \cdot 0{,}25 = 0$$
$$E(Y) = 0 \cdot 0{,}5 + 1 \cdot 0{,}5 = 0{,}5$$

El siguiente paso consiste en calcular los productos $X \cdot Y$ y calcular su esperanza. Téngase en cuenta que estos productos solo pueden tomar valores –1, 0,+1 y sus probabilidades vendrán dadas por:

$$P(X \cdot Y = 0) = P(X = 0 \text{ o } Y = 0) = P(X = 0) + P(Y = 0) - P(X = 0, Y = 0) =$$
$$= 0{,}50 + 0{,}50 - 0{,}50 = 0{,}50$$

$P(X \cdot Y = -1) = P(X = -1, Y = 1) = 0{,}25$ (directamente de la tabla)

$P(X \cdot Y = 1) = P(X = 1, Y = 1) = 0{,}25$ (directamente de la tabla)

A partir de estos cálculos es inmediato ver que $E(X \cdot Y) = 0$ y que cov(X, Y) = 0.

b. Cuando la covarianza entre dos variables aleatorias es 0, las variables pueden ser independientes o dependientes. El dato calculado no nos permite, en este caso, discernirlo. Necesitamos usar otro tipo de enfoque. Podemos ver que las variables son dependientes pues, por ejemplo, es fácil ver que:

$$P(X = 0, Y = 0) = 0{,}5 \neq P(X = 0) \cdot P(Y = 0) = 0{,}5 \cdot 0{,}5$$

y las variables no pueden ser independientes ya que, al menos en un caso, no se cumple la regla multiplicativa de la probabilidad.

92. Considere dos experimentos aleatorios. En el primero lanzamos al aire dos monedas honestas. En el segundo extraemos al azar una bola de una urna que contiene tres bolas numeradas del 1 al 3. Considere las variables aleatorias X, que mide el número de caras en el primer experimento, e Y, que mide el número de la bola extraída en el segundo.

a. Obtener la función de probabilidad conjunta $f(x,y)$ del vector aleatorio (X,Y) y construir la matriz de probabilidades. Verificar que $\sum_{i,j} f(x_i, y_j) = 1$
b. A partir de la matriz de probabilidades recupere las funciones de probabilidad de X e Y.
c. Discutir si X e Y son independientes.
d. Considerar una nueva variable aleatoria $Z=X \cdot Y$. Obtener la distribución de probabilidad de Z y a partir de ella calcule $E(Z)$. ¿De qué otra manera hubiéramos podido calcular $E(Z)$?
e. Considerar ahora $Z=X+Y$. Obtener la distribución de probabilidad de Z y a partir de ella calcule var(Z). ¿De qué otra manera hubiéramos podido calcular var(Z)?

a. Construimos en primer lugar una tabla con todos los posibles resultados de los dos experimentos. En cada una de las entradas de la tabla introduciremos los valores correspondientes a las variables X (cara: +, cruz: –) e Y (1, 2, 3):

Monedas / Bola	1	2	3
$--$	(0,1)	(0,2)	(0,3)
$-+$	(1,1)	(1,2)	(1,3)
$+-$	(1,1)	(1,2)	(1,3)
$++$	(2,1)	(2,2)	(2,3)

Dado que las monedas son honestas y cada bola tiene la misma probabilidad de ser extraída, tenemos que todas las casillas de esta tabla son equiprobables. En otras palabras, cada uno de los 12 posibles resultados de la combinación de los 2 experimentos tiene la misma probabilidad de ocurrir: $1/12$.

Construimos a continuación la matriz de probabilidades de la variable bidimensional (X, Y) a partir de los datos de la tabla anterior. Para determinar la probabilidad de cada par de valores (x, y) simplemente miraremos en cuántos resultados aparece ese par de valores y multiplicaremos ese número por $1/12$, que es la probabilidad de cada resultado. De esta manera obtenemos:

X / Y	1	2	3
0	$1/12$	$1/12$	$1/12$
1	$2/12$	$2/12$	$2/12$
2	$1/12$	$1/12$	$1/12$

Claramente, comprobamos que $\sum_{i,j} f(x_i, y_j) = 1$

b. Hallamos ahora las funciones de probabilidad marginal de X e Y, f_X y f_Y:

$X \setminus Y$	1	2	3	f_X
0	$1/12$	$1/12$	$1/12$	$1/4$
1	$2/12$	$2/12$	$2/12$	$1/2$
2	$1/12$	$1/12$	$1/12$	$1/4$
f_Y	$1/3$	$1/3$	$1/3$	1

c. Para comprobar si X e Y son independientes es necesario comprobar si para cada par de valores (x,y) se cumple la condición de independencia:

$$f_{X \cdot Y}(x,y) = f_X(x) \cdot f_Y(y)$$

condición que claramente se cumple en este caso si observamos los valores de la matriz de probabilidades conjuntas junto con las marginales.

Intuitivamente, este resultado de independencia entre X e Y no debe sorprender: el hecho que salgan más o menos caras al lanzar las monedas no tiene nada que ver con la bola que extraeremos de la urna.

d. La nueva variable Z puede tomar valores en $\{0,1,2,3,4,6\}$ con probabilidades:

$$P(Z=0) = P(X=0, Y=1) + P(X=0, Y=2) + P(X=0, Y=3) = \frac{3}{12}$$

$$P(Z=1) = P(X=1, Y=1) = \frac{2}{12}$$

$$P(Z=2) = P(X=1, Y=2) + P(X=2, Y=1) = \frac{3}{12}$$

$$P(Z=3) = P(X=1, Y=3) = \frac{2}{12}$$

$$P(Z=4) = P(X=2, Y=2) = \frac{1}{12}$$

$$P(Z=6) = P(X=2, Y=3) = \frac{1}{12}$$

Por tanto,

$$E(Z) = 0 \cdot \frac{3}{12} + 1 \cdot \frac{2}{12} + 2 \cdot \frac{3}{12} + 3 \cdot \frac{2}{12} + 4 \cdot \frac{1}{12} + 6 \cdot \frac{1}{12} = 2$$

Dado que hemos visto que X e Y son independientes, una forma alternativa de calcular $E(Z)$ hubiera sido:

$$E(Z) = E(X \cdot Y) = E(X) \cdot E(Y)$$

donde,

$$E(X) = 0 \cdot \frac{3}{12} + 1 \cdot \frac{6}{12} + 2 \cdot \frac{3}{12} = 1$$
$$E(Y) = 1 \cdot \frac{4}{12} + 2 \cdot \frac{4}{12} + 3 \cdot \frac{4}{12} = 2$$

Por tanto,

$$E(Z) = E(X \cdot Y) = E(X) \cdot E(Y) = 2 \cdot 1 = 2$$

e. La nueva variable $Z=X+Y$ puede tomar valores en $\{1,2,3,4,5\}$ con probabilidades:

$$P(Z=1) = P(X=0, Y=1) = \frac{1}{12}$$
$$P(Z=2) = P(X=0, Y=2) + P(X=1, Y=1) = \frac{3}{12}$$
$$P(Z=3) = P(X=0, Y=3) + P(X=1, Y=2) + P(X=2, Y=1) = \frac{4}{12}$$
$$P(Z=4) = P(X=1, Y=3) + P(X=2, Y=2) = \frac{3}{12}$$
$$P(Z=5) = P(X=2, Y=3) = \frac{1}{12}$$

Por tanto,

$$E(Z) = 1 \cdot \frac{1}{12} + 2 \cdot \frac{3}{12} + 3 \cdot \frac{4}{12} + 4 \cdot \frac{3}{12} + 5 \cdot \frac{1}{12} = 3$$
$$\mathrm{var}(Z) = 1^2 \cdot \frac{1}{12} + 2^2 \cdot \frac{3}{12} + 3^2 \cdot \frac{4}{12} + 4^2 \cdot \frac{3}{12} + 5^2 \cdot \frac{1}{12} - (3)^2 = \frac{122}{12} - 9 = \frac{7}{6}$$

Dado que hemos visto que X e Y son independientes, una forma alternativa de calcular var(Z) hubiera sido:

$$\mathrm{var}(Z) = \mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$$

donde,

$$\mathrm{var}(X) = 0^2 \cdot \frac{3}{12} + 1^2 \cdot \frac{6}{12} + 2^2 \cdot \frac{3}{12} - (1)^2 = \frac{18}{12} - 1 = \frac{3}{6}$$
$$\mathrm{var}(Y) = 1^2 \cdot \frac{4}{12} + 2^2 \cdot \frac{4}{12} + 3^2 \cdot \frac{4}{12} - (2)^2 = \frac{56}{12} - 4 = \frac{4}{6}$$

Por tanto,

$$\mathrm{var}(Z) = \mathrm{var}(X+Y) = \mathrm{var}(X) + V(Y) = \frac{3}{6} + \frac{4}{6} = \frac{7}{6}$$

93. (*) La variable aleatoria X representa el número de preguntas contestadas correctamente en un examen de Estadística I compuesto por tres preguntes. La variable Y representa el número de preguntas contestadas correctamente en un examen de Microeconomía I que solo tiene dos preguntas. La matriz de probabilidades conjuntas del vector aleatorio (X, Y) está descrita en la siguiente tabla:

Y/X	*0*	*1*	*2*	*3*
0	0,20	0,03	0,02	0
1	0	0,30	0,10	0,10
2	0	0,01	0,04	0,20

a. Averiguar cuál es el número esperado de respuestas correctas en el examen de Estadística I, $E(X)$, y de Micro I, $E(Y)$.
b. Valorar la validez de la siguiente afirmación: “Los alumnos que responden bien en el examen de Estadística, también se espera que respondan bien en el de Micro”.
c. Consideremos ahora el grupo de estudiantes que han contestado correctamente las tres preguntas del examen de Estadística. Para este grupo, ¿cuál es el número esperado de respuestas correctas en el examen de Micro?

a. En primer lugar necesitamos calcular las distribuciones marginales de probabilidad para la variable X (sumando columnas) y para la variable Y (sumando filas):

$P(X=0) = 0{,}20$, $P(X=1) = 0{,}34$, $P(X=2) = 0{,}16$, $P(X=3) = 0{,}30$

$P(Y=0) = 0{,}25$, $P(Y=1) = 0{,}50$, $P(Y=2) = 0{,}25$

A partir de esta información calculamos las esperanzas respectivas y obtenemos:

$E(X) = 1{,}56$, $E(Y) = 1$

b. Hemos de ver si X e Y tienen alguna conexión estadística entre si. Una manera de verlo es calculando la covarianza entre X e Y. Recordemos que si $\text{cov}(X,Y) > 0$, entonces las variables X e Y tienen un vínculo directo y si $\text{cov}(X,Y) < 0$, entonces presentan un vínculo inverso. Podemos calcular la covarianza usando esta expresión:

$$\text{cov}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y).$$

Puesto que anteriormente hemos calculado las esperanzas de X y de Y solo necesitamos ahora calcular la esperanza del producto:

$$E(X\cdot Y) = (0\cdot 0)\cdot P(X{=}0 \text{ y } Y{=}0) + (0\cdot 1)\cdot P(X{=}0 \text{ y } Y{=}1) + \ldots.$$
$$+ (2\cdot 3)\cdot P(X{=}2 \text{ y } Y{=}3) = 2{,}18$$

Obtenemos pues que:

$$\text{cov}(X, Y) = 2{,}18 - (1{,}56\cdot 1) = 0{,}62$$

Al tener covarianza positiva podemos concluir que los alumnos que responden favorablemente en el examen de Estadística también tenderán a responder positivamente en el de Microeconomía.

c. Hemos de calcular la esperanza de las respuestas correctas en Micro I bajo la condición que nos fijamos en los estudiantes que han contestado bien las 3 preguntas de Estadística. Hemos de ir a la tercera columna de la matriz de probabilidades y calcular la esperanza que corresponde a la variable Y sujeta a la condición que $X = 3$:

$$E(Y / X{=}3) = \sum_{j=0}^{2} y_j \cdot P(Y = y_j / X = 3)$$

Recordando la definición de probabilidad condicional:

$$E(Y/X{=}3) = \sum_{j=0}^{2} y_j \cdot P(Y = y_j / X = 3) = \sum_{j=0}^{2} y_j \cdot \frac{P(Y = y_j \cap X = 3)}{P(X = 3)}$$

En las fracciones de la expresión las probabilidades del numerador se toman directamente de la matriz mientras que la del denominador se ha calculado en el apartado *a*) anterior. Por consiguiente:

$$E(Y/X{=}3) = 0 \cdot \frac{P(Y = 0 \cap X = 3)}{P(X = 3)} + 1 \cdot \frac{P(Y = 1 \cap X = 3)}{P(X = 3)} + 2 \cdot \frac{P(Y = 2 \cap X = 3)}{P(X = 3)} =$$

$$= 0 \cdot \frac{0}{0{,}3} + 1 \cdot \frac{0{,}1}{0{,}3} + 2 \cdot \frac{0{,}2}{0{,}3} = 1{,}667$$

94. (*) Vamos a considerar una variable aleatoria X que indica el número de años que los estudiantes de la Facultad de Economía y Empresa de la UAB están en paro tras finalizar sus estudios. Se sabe que como máximo tardan 1 año en encontrar trabajo, por ello X=0,1. Sea ahora Y la variable aleatoria que mide el número de respuestas correctas que estos mismos estudiantes tuvieron en el examen de Estadística I, con posibles valores Y=0,1,2,3,4. Se han tabulado las proporciones del par (X, Y):

Y/X	*0*	*1*
0	0,01	0,28
1	0,07	0,17
2	0,13	0,08
3	0,10	0,01
4	0,15	0,00

a. Explicar si esta tabla representa una distribución conjunta de probabilidad y, en ese caso, encontrar las distribuciones marginales de probabilidad.
b. Calcular la covarianza entre X e Y.
c. Explicar si la siguiente afirmación es verdadera o falsa: “Hay alguna cosa que sistemáticamente pasa entre el rendimiento académico y el tiempo en el paro y se puede afirmar que los alumnos con mejor rendimiento en los exámenes están menos tiempo sin trabajo.”

a. Se debe verificar si la suma de las proporciones de las 10 celdas de la tabla es la unidad (hágalo y verá que lo es). Las distribuciones marginales se calculan sumando las proporciones de las filas (para la de *Y*) y las de las columnas (para *X*). Hágalo y anote los resultados.

b. La covarianza se obtiene de: $\text{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$.
De las distribuciones marginales hallaríamos $E(X)$ y $E(Y)$. Por ejemplo:

$$E(X) = 0 \cdot 0,46 + 1 \cdot 0,54 = 0,54$$

También encontraríamos que $E(Y)$=1,59. Para obtener $E(X \cdot Y)$ debemos construir todos los posibles productos entre *X* e *Y* y ver en cada una de las celdas cuál es la probabilidad conjunta de cada uno de los pares. Así:

$$E(X \cdot Y) = (0 \cdot 0) \cdot P(X = 0, Y = 0) + (0 \cdot 1) \cdot P(X = 0, Y = 1) + ...$$
$$+(4 \cdot 1) \cdot P(X = 4, Y = 1) = 0,36$$

Finalmente,

$$\text{cov}(X,Y) = 0,36 - 0,54 \cdot 1,59 = -0,49$$

c. Puesto que la covarianza es diferente de 0 es seguro que existe una relación estadística de dependencia entre *X* e *Y*; en otras palabras, existe un cierto vínculo entre rendimiento académico y tiempo en el paro y además este vínculo es negativo: a mayor rendimiento en los exámenes, menor es el tiempo en el paro. En consecuencia podemos concluir que la afirmación es verdadera.

95. Sean *X* e *Y* dos variables aleatorias discretas con función de probabilidad conjunta dada en la siguiente tabla:

X/*Y*	0	1
0	0	2/6
1	2/6	1/6
2	1/6	0

a. Obtener las funciones de probabilidad marginales de las dos variables, f_X e f_Y.
b. Si nos dicen que la variable *Y* ha tomado el valor 0, ¿cuál es la probabilidad de que *X* tome el valor 1?
c. Comprobar si las dos variables son estocásticamente independientes.
d. Calcular la covarianza entre las dos variables, cov(*X*, *Y*).

a. La tabla siguiente muestra los resultados:

X/Y	0	1	f_X
0	0	2/6	**2/6**
1	2/6	1/6	**3/6**
2	1/6	0	**1/6**
f_Y	**3/6**	**3/6**	1

b. Calculando la probabilidad condicionada obtenemos:

$$P(X=1/Y=0)=f_{X/Y=0}(1)=\frac{f_{XY}(1,0)}{f_Y(0)}=\frac{2/6}{3/6}=\frac{2}{3}$$

c. Para que dos variables X e Y sean independientes se debe cumplir que:

$$f_{XY}(x,y)=f_X(x)\cdot f_Y(y)\quad \forall (x,y)\in Supp(X,Y)$$

donde *Supp*(X, Y) representa todos los posibles pares de valores construibles con los valores de X y de Y. En este caso podemos comprobar que para los valores (0,0) esto **NO** se cumple:

$$f_{XY}(0,0)=0\neq\frac{2}{6}\cdot\frac{3}{6}\,f_X(0)\cdot f_Y(0)$$

Por tanto, en este caso X e Y **NO** son independientes:

d. Para calcular la covarianza entre X e Y tenemos:

$$\operatorname{cov}(X,Y)=E(X\cdot Y)-E(X)\cdot E(Y)$$

donde,

$$E(X\cdot Y)=\sum_{(x,y)\in Supp(X,Y)} x\cdot y\cdot f_{XY}(X,Y)=0\cdot0\cdot0+0\cdot1\cdot\frac{2}{6}+1\cdot0\cdot\frac{2}{6}+1\cdot1\cdot\frac{1}{6}+2\cdot0\cdot\frac{1}{6}+2\cdot1\cdot0=\frac{1}{6}$$

$$E(X)=\sum_{x\in Supp(X)} x\cdot f_X(x)=0\cdot\frac{2}{6}+1\cdot\frac{3}{6}+2\cdot\frac{1}{6}=\frac{5}{6}$$

$$E(Y)=\sum_{y\in Supp(Y)} y\cdot f_Y(y)=0\cdot\frac{3}{6}+1\cdot\frac{3}{6}=\frac{3}{6}$$

Por tanto,

$$\text{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = \frac{1}{6} - \frac{5}{6} \cdot \frac{3}{6} = -\frac{1}{4}$$

96. (*) Considere una variable aleatoria X que toma valores –1, 0, +1 con probabilidades 1/4, 1/2, 1/4, respectivamente. Defina una nueva variable aleatoria Y mediante $Y = X^2$.

   a. Usando la definición de Y obtener la función de probabilidad de Y.
   b. Construir la matriz de probabilidades conjuntas del vector aleatorio (X, Y).
   c. A partir de la matriz de probabilidades, obtener las funciones marginales de probabilidad de X y de Y. Comparar las marginales con la distribución dada de X y la derivada para Y en el apartado *a*.
   d. Discutir si X e Y son variables aleatorias independientes: i) usando la definición de independencia entre variables aleatorias, ii) usando el sentido común.
   e. Calcular la covarianza de X e Y y discutir su valor en relación a la respuesta del apartado anterior.
   f. Calcular el coeficiente de correlación entre X e Y. Discutir e interpretar a la luz de todo lo anterior.

a. Claramente, la nueva variable Y tan solo puede tomar valores en el conjunto $\{1,2\}$ de forma que Y=0 si X=0, e Y=1 si X=-1 o si X=1. Para cada uno de estos valores tenemos:

$$P(Y=0) = P(X=0) = \frac{1}{2}$$

$$P(Y=1) = P(X=-1 \text{ o } X=1) = P(X=-1) + P(X=1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Por tanto, la función de probabilidad de la nueva variable Y será:

$$f_Y(y) = \begin{array}{ll} 1/2 & y \in \{0,1\} \\ 0 & \text{en caso contrario} \end{array}$$

b. Para obtener la función de probabilidad conjunta utilizaremos la relación que se obtiene de la fórmula básica de la probabilidad condicionada:

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(B/A) \cdot P(A)$$

Aplicado al caso que nos ocupa tendremos:

$$P(X = x \cap Y = y) = P(Y = y / X = x) \cdot P(X = x)$$

Así, para cada una de las posibles combinaciones de valores de X y de Y tenemos:

$$P(X = -1 \cap Y = 0) = P(Y = 0 / X = -1) \cdot P(X = -1) = 0 \cdot \frac{1}{4} = 0$$
$$P(X = -1 \cap Y = 1) = P(Y = 1 / X = -1) \cdot P(X = -1) = 1 \cdot \frac{1}{4} = \frac{1}{4}$$
$$P(X = 0 \cap Y = 0) = P(Y = 0 / X = 0) \cdot P(X = 0) = 1 \cdot \frac{1}{2} = \frac{1}{2}$$
$$P(X = 0 \cap Y = 1) = P(Y = 1 / X = 0) \cdot P(X = 0) = 0 \cdot \frac{1}{2} = 0$$
$$P(X = 1 \cap Y = 0) = P(Y = 0 / X = 1) \cdot P(X = 1) = 1 \cdot \frac{1}{4} = 0$$
$$P(X = 1 \cap Y = 1) = P(Y = 1 / X = 1) \cdot P(X = 1) = 1 \cdot \frac{1}{4} = \frac{1}{4}$$

lo que nos genera la matriz de probabilidades siguiente:

X / Y	0	1
-1	0	$\frac{1}{4}$
0	$\frac{1}{2}$	0
1	0	$\frac{1}{4}$

c. Obtenemos las funciones de probabilidad marginal de X e Y sumando por filas y columnas.

d.

X/Y	0	1	f_X
-1	0	$1/4$	$1/4$
0	$1/2$	0	$1/2$
1	0	$1/4$	$1/4$
f_Y	$1/2$	$1/2$	1

Comprobamos que las funciones de probabilidad marginal coinciden con las que teníamos anteriormente: f_X coincide con las probabilidades dadas en el enunciado del ejercicio y f_Y coincide con las probabilidades obtenidas en el apartado *b* del ejercicio.

e. Estudiaremos ahora la independencia entre X e Y
i) Según la definición de independencia estocástica, X e Y serán independientes si y solo si para todo par de valores (x, y) se cumple que:

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

Podemos comprobar rápidamente que esto no se cumple. Efectivamente, tomando el par $(x, y) = (-1,0)$ vemos que $f_{XY}(-1,0) = 0$ mientras que:

$$f_X(-1) \cdot f_Y(0) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8} \neq 0$$

Por tanto, no se cumple la condición de independencia entre las variables.

ii) Utilizando el "sentido común", es evidente que existe una clara relación entre las variables. De hecho, la relación es total ya que el valor de la variable Y está totalmente determinado por el valor de la variable X por medio de la relación

$$Y = X^2$$

f. Obtendremos la covarianza entre X e Y, cov(X,Y), mediante la expresión:

$$\text{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Calculamos en primer lugar las esperanzas:

$$E(X) = (-1)\frac{1}{4} + 0\frac{1}{2} + (1)\frac{1}{4} = 0$$
$$E(Y) = 0\frac{1}{2} + 1\frac{1}{2} = \frac{1}{2}$$
$$E(X \cdot Y) = (-1)\cdot 0 \cdot 0 + (-1)\cdot 1 \cdot \frac{1}{4} + 0 \cdot 0 \cdot \frac{1}{2} +$$
$$+0 \cdot 1 \cdot 0 + 1 \cdot 0 \cdot 0 + 1 \cdot 1 \cdot \frac{1}{4} = -\frac{1}{4} + \frac{1}{4} = 0$$

Por tanto,

$$\operatorname{cov}(X,Y) = E(X \cdot Y) - E(X)\cdot E(Y) = 0 - 0\frac{1}{2} = 0$$

La covarianza entre las dos variables resulta ser <u>cero</u>.

Este resultado ilustra el hecho que, si bien dos variables independientes necesariamente tienen una covarianza igual a cero, la afirmación contraria no es cierta en general. Es decir, una covarianza igual a cero no es suficiente para asegurar que las variables en cuestión sean independientes. Esquemáticamente,

$$X,Y \text{ independientes} \Rightarrow \operatorname{cov}(X,Y) = 0$$
$$\operatorname{cov}(X,Y) = 0 \not\Rightarrow X,Y \text{ independientes}$$

g. Dado que el coeficiente de correlación, ρ_{XY}, se obtiene a partir de:

$$\rho_{XY} = \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(X)\cdot \operatorname{var}(Y)}}$$

y que hemos obtenido $\operatorname{cov}(X,Y) = 0$, tendremos que:

$$\rho_{XY} = 0$$

si bien las dos variables están claramente relacionadas.

Las razones de este resultado son las mismas que en el apartado anterior: la covarianza y, en consecuencia, el coeficiente de correlación pueden resultar igual a cero aún existiendo una relación entre las variables.

97. Ver que si $Y = -a \cdot X$, donde $a>0$, entonces el coeficiente de correlación entre X e Y es -1.

Recordemos en primer lugar que el coeficiente de correlación entre X e Y, ρ_{XY} , se obtiene a partir de la expresión:

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

donde $\mathrm{cov}(X,Y)$ representa la covarianza entre X e Y, y σ_X y σ_Y las desviaciones típicas de X e Y respectivamente. Para verificar el enunciado propuesto trabajaremos con las propiedades de la esperanza, la varianza y la covarianza. Así, por ejemplo, tenemos que:

$$E(Y) = E(-a \cdot X) = -a \cdot E(X)$$
$$\mathrm{var}(Y) = \mathrm{var}(-a \cdot X) = a^2 \cdot \mathrm{var}(X) \Rightarrow \sigma_Y^2 = a^2 \cdot \sigma_X^2 \Rightarrow \sigma_Y = a \;\; \cdot \sigma_X$$

y de la misma manera,

$$E(X \cdot Y) = E(X(-a \cdot X)) = E(-a \cdot X^2) = -a \cdot E(X^2)$$

Calculando ahora la covarianza ente X e Y tenemos:

$$\mathrm{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) =$$
$$= -a \cdot E(X^2) - E(X) \cdot (-a \cdot E(X)) =$$
$$= -a \cdot (E(X^2) + a \cdot E(X)^2 =$$
$$= -a \cdot [E(X^2) - E(X)^2] = -a \cdot \sigma_X^2$$

ya que, por definición de varianza:

$$\mathrm{var}(X) = E[X - E(X)]^2 = E(X^2) - E(X)^2$$

Por tanto, combinando los resultados obtenidos vemos que:

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{-a \cdot \sigma_X^2}{\sigma_X \cdot (a \cdot \sigma_X)} = \frac{-a \cdot \sigma_X^2}{a \cdot \sigma_X^2} = -1$$

tal como queríamos demostrar.

98. Ver que si $Y=aX+b$, donde $a>0$, entonces el coeficiente de correlación entre X e Y es 1.

De forma similar al ejercicio anterior tenemos que:

$$E(Y) = E(a \cdot X + b) = a \cdot E(X) + b$$
$$\text{var}(Y) = \text{var}(a \cdot X + b) = a^2 \cdot \text{var}(X) + 0 \Rightarrow \sigma_Y^2 = a^2 \cdot \sigma_X^2 \Rightarrow \sigma_Y = a \cdot \sigma_X$$

y de la misma manera,

$$E(X \cdot Y) = E(X(a \cdot X + b)) = E(a \cdot X^2 + b \cdot X) = a \cdot E(X^2) + b \cdot E(X)$$

Calculando ahora la covarianza ente X e Y tenemos:

$$\begin{aligned}
&\text{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = \\
&= a \cdot E(X^2) + b \cdot E(X) - E(X) \cdot (a \cdot E(X) + b) = \\
&= a \cdot E(X^2) + b \cdot E(X) - a \cdot E(X)^2 - b \cdot E(X) = \\
&= a \cdot [E(X^2) - E(X)^2] = a \cdot \sigma_X^2
\end{aligned}$$

ya que, de nuevo por definición de varianza:

$$\text{var}(X) = E[X - E(X)]^2 = E(X^2) - E(X)^2$$

Por tanto, combinando los resultados obtenidos vemos que:

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{a \cdot \sigma_X^2}{\sigma_X \cdot (a \cdot \sigma_X)} = \frac{a \cdot \sigma_X^2}{a \cdot \sigma_X^2} = 1$$

tal como queríamos demostrar.

99. Un investigador quiere saber si al anunciar un nuevo producto por TV, este acaba siendo más conocido por las personas que ven regularmente la TV a la hora de emitirse el anuncio que por los que no la ven. Suponer que el anuncio se emite en el intermedio del partido de fútbol semanal. Después de tabular los datos se sabe que las personas que ven regularmente el programa y reconocen el nuevo producto son el 15%. Además, el programa es visto por un 16% de los televidentes encuestados y el nuevo producto es reconocido por el 45% de los encuestados. Considere dos variables aleatorias X e Y tales que: X=1 si un encuestado ve el programa y X=0 si no lo ve; Y=1 si reconoce el producto, e Y=0 en caso contrario.

a. Construir la matriz de probabilidades del par (X, Y).
b. Hallar la función de probabilidad condicional de Y para ambos valores de X.
c. Hallar e interpretar la covarianza entre X e Y.

a. Leyendo e interpretando la información proporcionada podemos construir fácilmente, de forma parcial de momento, la siguiente tabla de probabilidades:

X/Y	0	1	f_X
0			
1		0,15	0,16
f_Y		0,45	

A partir de aquí podemos completar la información que falta utilizando las propiedades de las probabilidades. Así, sabiendo las proporciones de los que ven el programa y de los que reconocen el producto, podemos completar en primer lugar las proporciones de los que NO ven el programa y de los que NO reconocen el producto (ya que en cada caso han de sumar 1).

X/Y	0	1	f_X
0			0,84
1		0,15	0,16
f_Y	0,55	0,45	1

y a partir de aquí podemos ir completando el resto hasta obtener:

X/Y	0	1	f_X
0	0,54	0.30	0,84
1	0,01	0,15	0,16
f_Y	0,55	0,45	1

b. Nos piden ahora hallar las funciones $f_{Y/X=0}(x)$ y $f_{Y/X=1}(x)$. Recordemos que:

$$f_{Y/X=0} = \frac{f_{XY}(0,y)}{f_X(0)} = \frac{f_{XY}(0,y)}{0,84}$$

$$f_{Y/X=1} = \frac{f_{XY}(1,y)}{f_X(1)} = \frac{f_{XY}(1,y)}{0,16}$$

Así obtenemos:

$$f_{Y/X=0}(0) = \frac{f_{XY}(0,0)}{0,84} = \frac{0,54}{0,84} = 0.643 \qquad f_{Y/X=1}(0) = \frac{f_{XY}(1,0)}{0,16} = \frac{0,01}{0,16} = 0.0625$$

$$f_{Y/X=0}(1) = \frac{f_{XY}(0,1)}{0,84} = \frac{0,30}{0,84} = 0,357 \qquad f_{Y/X=1}(1) = \frac{f_{XY}(1,1)}{0,16} = \frac{0,15}{0,16} = 0,9375$$

c. Finalmente, la covarianza entre X e Y la obtenemos mediante la expresión:

$$\operatorname{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Donde,

$$E(X \cdot Y) = \sum_{x=0}^{1}\sum_{y=0}^{1} x \cdot y \cdot f_{XY}(x,y) =$$

$$= 0 \cdot 0 \cdot 0,54 + 0 \cdot 1 \cdot 0,30 + 1 \cdot 0 \cdot 0,01 + 1 \cdot 1 \cdot 0,15 = 0,15$$

y

$$E(X) = \sum_{x=0}^{1} x \cdot f_X(x) = 0 \cdot 0,84 + 1 \cdot 0,16 = 0,16$$

$$E(Y) = \sum_{y=0}^{1} y \cdot f_Y(y) = 0 \cdot 0,55 + 1 \cdot 0,45 = 0,45$$

Por lo tanto,

$$\mathrm{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0{,}15 - 0{,}16 \cdot 0{,}45 = 0{,}078$$

100. (*) El Consejo de Estudiantes tiene la sospecha que los profesores aparecen lo viernes por su despacho menos que en el resto de días de la semana. Para estudiar la situación deciden montar un sistema de control consistente en establecer un sistema de visitas aleatorias a los despachos de los profesores. El sistema establecido de visitas se repite religiosamente cada día de la semana durante un mes. Al final se sabe que en el 16% de las visitas realizadas los viernes no se encuentra al profesor en su despacho; esta cifra es del 12% para el resto de días de la semana. Estudiar este problema a fin de determinar hasta qué punto existe (o no) una relación entre el día de la semana y el absentismo.

a. Empiece por obtener la matriz de probabilidades conjuntas de (X, Y), donde X=1 si la visita se realiza en viernes y 0 en caso contrario, y donde Y=1 si el profesor no está en su despacho y 0 en caso contrario.
b. Obtener las funciones marginales de probabilidad.
c. Obtener las funciones de probabilidad condicional.
d. Obtener la covarianza y el coeficiente de correlación entre X e Y. Comentar.

a. La construcción de la matriz de probabilidades conjuntas requiere tener en cuenta los siguientes aspectos. En primer lugar distinguiremos las visitas efectuadas los viernes de las efectuadas el resto de días de la semana (no viernes), y las primeras representan el 20% de las visitas mientras que las segundas son el 80%. Esta información nos permite tener la suma de las columnas de la matriz de probabilidades si en ellas representamos a la variable X. Sabemos, además, los porcentajes de absentismo. Toda esta información nos permite tener estos datos:

	X=0	X=1	Marginal de Y
Y=0	?	?	?
Y=1	0,12	0,16	0,28
Marginal de X	0,80	0,20	1

Las celdas restantes se obtienen por diferencia teniendo en cuenta que el total de probabilidad ha de ser necesariamente 1:

	X=0	X=1	Marginal de Y
Y=0	0,68	0,04	0,72
Y=1	0,12	0,16	0,28
Marginal de X	0,80	0,20	1

b. Directamente podemos ver que:

$$P(X=0)=0{,}80$$
$$P(X=1)=0{,}20$$
$$P(Y=0)=0{,}72$$
$$P(Y=1)=0{,}28$$

c. A título de ejemplo, si queremos calcular la probabilidad que el profesor no esté en su despacho ($Y=1$) dado que es viernes ($X=1$) podemos partir de:

$$P(X=1\cap Y=1)=P(X=1)\cdot P(Y=1/X=1)=0{,}16$$

y de aquí,

$$P(Y=1/X=1)=\frac{P(X=1\cap Y=1)}{P(X=1)}=\frac{0{,}16}{0{,}20}=80\%$$

Similarmente, la probabilidad que el profesor esté ausente si no es viernes se evaluará por:

$$P(Y=1/X=0)=\frac{P(X=0\cap Y=1)}{P(X=0)}=\frac{0{,}12}{0{,}80}=15\%$$

Claramente, los viernes son un mal día si los estudiantes desean ver a su profesor.

d. Calcularemos la covarianza entre X y Y a partir de la expresión:

$$\text{cov}(X,Y)=E(X\cdot Y)-E(X)\cdot E(Y)$$

En primer lugar es fácil ver a partir de la definición de esperanza y los datos de la matriz de probabilidades que las esperanzas de X y de Y son:

$$E(X)=0\cdot 0{,}80+1\cdot 0{,}20=0{,}20$$
$$E(Y)=0\cdot 0{,}72+1\cdot 0{,}28=0{,}28$$

En segundo lugar, podemos ver que el producto $X \cdot Y$ solo puede tomar dos valores, a saber, $X \cdot Y = 0$ y $X \cdot Y = 1$. Estos sucesos se darán en las siguientes circunstancias:

$X \cdot Y = 0$ si y solo si $X = 0$ o Y =0
$X \cdot Y = 1$ si y solo si $X = 1$ y $Y = 1$

Hallamos las probabilidades de estos sucesos:

$$P(X \cdot Y = 0) = P(X = 0 \cup Y = 0) =$$
$$= P(X = 0) + P(Y = 0) - P(X = 0 \cap Y = 0) = 0{,}80 + 0{,}72 - 0{,}68 = 0{,}84$$

$$P(X \cdot Y = 1) = P(X = 1 \cap Y = 1) = 0{,}16$$

Podemos ahora calcular la esperanza del producto:

$$E(X \cdot Y) = 0 \cdot 0{,}84 + 1 \cdot 0{,}16 = 0{,}16$$

y finalmente la covarianza será:

$$\text{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0{,}16 - 0{,}20 \cdot 0{,}84 = 0{,}104$$

Para derivar el coeficiente de correlación entre X y Y necesitamos calcular en primera instancia sus varianzas. En el caso de la variable X y usando la expresión:

$$\text{var}(X) = E(X^2) - E(X)^2$$

podemos calcular $E(X^2)$ fácilmente:

$$E(X^2) = 0^2 \cdot 0{,}80 + 1^2 \cdot 0{,}20 = 0{,}20$$

de forma que:

$$\text{var}(X) = E(X^2) - E(X)^2 = 0{,}20 - 0{,}20^2 = 0{,}16$$

Procediendo de forma idéntica para la variable Y (dejamos el cálculo al lector) obtendríamos:

$$\text{var}(Y) = 0{,}2016$$

El coeficiente de correlación buscado será pues:

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\cdot\text{var}(Y)}} = \frac{0{,}104}{\sqrt{0{,}16\cdot 0{,}2016}} = 0{,}58$$

A la luz de los cálculos, podemos concluir que existe una correlación positiva entre el día de la semana y el grado de absentismo del profesorado.